A USER-CENTERED ANALYSIS OF VIRTUAL REALITY IN DESIGN REVIEW: COMPARING THREE-DIMENSIONAL PERCEPTION AND PRESENCE BETWEEN IMMERSIVE AND NON-IMMERSIVE ENVIRONMENTS

A Dissertation
Presented to
The Academic Faculty

by

Daniel Paes

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in Building Construction

Georgia Institute of Technology

December 2019

A USER-CENTERED ANALYSIS OF VIRTUAL REALITY IN DESIGN REVIEW: COMPARING THREE-DIMENSIONAL PERCEPTION AND PRESENCE BETWEEN IMMERSIVE AND NON-IMMERSIVE ENVIRONMENTS

Approved by:

Dr. Javier Irizarry, Advisor
School of Building Construction
Georgia Institute of Technology

Dr. Sonit Bafna, Minor Advisor
School of Architecture
Georgia Institute of Technology

Dr. Daniel Castro
School of Building Construction
Georgia Institute of Technology

Dr. Burcin Becerik-Gerber
Sonny Astani Department of Civil and Environmental Engineering
University of Southern California

Dr. Masoud Gheisari
M.E. Rinker, Sr. School of Construction Management
University of Florida

Date Approved: 12/04/2019

ACKNOWLEDGMENTS

Firstly, I would like to express my sincere gratitude to my advisor, Dr. Javier Irizarry, for his continued support, guidance and encouragement. I could not have asked for a better mentor to guide me through my Ph.D. journey. Working with him over these past years has been a truly edifying and life-changing experience.

I would also like to give special thanks to my minor advisor, Dr. Sonit Bafna, for his great contributions to this work, and to Professor Daniel Castro, for his continued support and advice. I am also deeply grateful to Professor Burcin Becerik-Gerber and Dr. Masoud Gheisari for their valuable insights and for accepting the invitation to take place in the Dissertation Committee. I wish to express my warm thanks to Dr. Diego Pujoni, for his enormous assistance with the statistical analyses. I also thank the entire faculty in the School of Building Construction and across Georgia Tech from whom I had the honor to learn so much. I would also like to thank the School's staff, especially Ms. Lorie Wooten and Ms. Laura Alger for their assistance, readiness and kindness. I warmly thank Gabriel Faura, Aline Valverde, and Gustavo do Amaral, whose friendship has made this journey a lot easier. I also owe gratitude to my fellow graduate students for the collaborations, for sharing their experiences, and for volunteering for my experiments. I thank CNPq for providing financial support (grant number 203359/2014-3).

Lastly, and most importantly, I thank my parents Ruy and Marisa, my sister Isabela, dearest Bruno, and Márcia for their love, for being there when I most needed, and for always believing in me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2D	Two-dimensional
3D	Three-dimensional
VR	Virtual Reality
VE	Virtual Environment
IE	Immersive Environment
IVR	Immersive Virtual Reality
niVR	Non-immersive Virtual Reality
PhE	Physical Environment
CVE	Collaborative Virtual Environment
CAD	Computer Aided Design
BIM	Building Information Modeling
HMD	Head-Mounted Display
HCI	Human-Computer Interaction
AECO-FM	Architecture, Engineering, Construction, Operation, and Facility Management
IRB	Institutional Review Board

SUMMARY

Over the last few years, the adoption of Virtual Reality (VR) solutions by the construction industry has grown rapidly worldwide. These have been developed and used for different purposes, including collaborative design review. Nonetheless, the extent to which such systems enhance the cognitive capabilities of construction professionals involved in the design review activity is still unclear. Knowledge on the cognitive benefits provided by Immersive Virtual Reality (IVR) technology is essential to elicit its usefulness and effectiveness, as well as to provide development directions. In this context, this study sought to quantitatively verify the ability of an IVR system in providing users with enhanced three-dimensional (3D) perception of a BIM (Building Information Modeling) model and greater levels of presence in the virtual environment (VE) compared to a non-immersive conventional VR system. The method compares users' 3D perception and levels of presence between two modes of presentation (IVR *vs.* non-immersive VR). The study also examines the relationship between 3D perception and presence within each virtual environment. Controlling for individual factors and order effects, findings indicate that in comparison to a conventional workstation, IVR technology improves 3D perception of the architectural model and provides more immersive experiences. Results also suggest no association between 3D perception and presence in virtual environments, contrary to expectations. The ability of IVR technology in providing current and future workforce with a significantly better understanding of the three-dimensional relationships of architectural models and greater levels of presence in the review task is expected to benefit collaborative design review.

CHAPTER 1

INTRODUCTION

The adoption of information technologies is an effective strategy to support, integrate, and optimize construction processes (Eastman et al., 2008). Advanced visualization solutions such as Immersive Virtual Reality (IVR) are implemented to improve decision-making and problem solving in collaborative design review and constructability analysis meetings (Maldovan et al., 2006; Okeil, 2010; Bassanimo et al., 2010; Dunston et al., 2011; Boton, 2018), in safety (Sacks et al., 2013) and disaster evacuation training (Lovreglio et al., 2018), in the prediction of human-building interactions (Bertol, 1997; Adi and Roberts, 2014; Heydarian et al., 2014, 2015a, 2015b, 2015c; Khashe et al., 2018), as well as in construction and architecture education (Castronovo et al., 2017; Lucas, 2018; Sopher et al., 2017, 2018).

In late 2014, the advent of powerful Virtual Reality (VR) software and devices – following recent advances in computer processing power, low latency tracking and display technologies – enabled the development and spread of VR solutions across the Architecture, Engineering, Construction, Operations and Facility Management (AECO-FM) industry, especially for design review purposes (Castronovo et al., 2017). The Startup Europe Partnership has identified VR as a key technology for Digital Construction and Industry 4.0 (SEP, 2018). The Hannover Fair has listed VR among the top twenty most potentially transformative technologies (IoT-Analytics, 2019). A recent report by Digi-Capital (2019) shows that the mixed-reality market is expected to reach US$90 billion in revenue by 2023. In 2018, the World Economic Forum along with the Boston Consulting Group (BCG, 2018) released a report in which mixed-reality is listed among the ten most promising technologies to improve productivity in construction. As per the KPMG Global Construction Survey (KPMG, 2016), visualization is the future of decision-making in capital projects.

Construction companies and design offices worldwide are exploring the benefits of IVR platforms for collaborative design review, including Gilbane, Mortenson Construction, and Perkins and Will. SHoP Architects (Shop Architects, 2019) is using IVR to supplement traditional representation techniques, improving client communication and speeding up the design process. The technology has allowed the design team at SHoP Architects to take bigger risks with clients in proposing complex spaces and effects otherwise hard to convey.

Typically, many stakeholders and experts from various disciplines participate in the design process. Design review meetings – also referred to as coordination meetings – take place at different stages of the design process, during which the participants get together to communicate, evaluate, merge, and generate design solutions (Yabuki, 2011). While the literature does not provide an exact description of a typical review meeting composition, case studies indicate that it is traditionally multi-disciplinary, although arrangements may vary across different project types and design phases (Eastman et al., 2008). For instance, while early (preliminary) design review can involve architects, owners and end-users, more technical meetings for clash detection and assessment of constructability issues in the detailed or construction design phases would involve designers, fabricators and contractors mostly (Yabuki, 2011). With the growing adoption of Concurrent Engineering, Lean Construction, and Integrated Project Delivery principles, an increasingly diverse set of collaborators will become involved in the early design stages (Eastman et al., 2008). Ultimately, a design review team can include owners, designers (architects and engineers), contractors and subcontractors, fabricators, construction managers, facility managers, technical specialist consultants, governmental officers, and even local community representatives (Eastman et al., 2008; Yabuki, 2011).

As values, knowledge, and terminology may vary among stakeholders, several mistakes can arise from miscommunication and misunderstanding in the collaborative design process (Yabuki, 2011). Seeking input from all stakeholders is difficult if they cannot adequately interpret the design representation at the review meetings (Eastman et al., 2008). In this context, innovative technology solutions such as VR-based Collaborative Virtual Environments (CVE) are expected to facilitate the communication and shared understanding of a project's 3D model (Yabuki, 2011; Tizani, 2011; Maher, 2011). The rich, interactive, and intuitive nature of 3D models can greatly enhance the participants' understanding of the design under revision (Eastman et al., 2008).

Studies on VR often rely on the definition coined by Steuer (1992), who strategically defined VR as a kind of presence experience, allowing for the distinction among different VR systems according to the level of presence provided. For Wann and Mon-Williams (1996), in order for a system to be considered a Virtual Environment (VE) it must satisfy criteria that arise from human spatial perception. Both Steuer's and Wann and Mon-Williams's definitions do not rely on a system's technological apparatus and appear better aligned with the general purpose of virtual environments in construction, i.e., to provide "realistic" experiences. In the context of design review, VR could be defined as the experience of presence in a fictitious architectural environment by means of its representation, whichever the type of representation in use (not necessarily digital three-dimensional models).

A key aspect in the design review activity is to understand the architectural design. Furthermore, a satisfactory and shared understanding of building representations throughout a project's life cycle is critical to ensure the success of construction projects. In this context, the representation method utilized to convey building information can largely affect the understanding of that information and the overall quality of decision-making and problem solving (Bassanimo et al., 2010). An adequate representation would communicate the designer's intentions in a less cognitively demanding format, allowing for a more effective and smooth review process (Castronovo et al., 2017). Because decisions of greater impact on costs, speed and quality of a project are made in the design phase, where the ability to influence the overall quality of construction is higher (Eastman et al., 2008; Owen et al., 2010), positive impacts such as decreased time and cost can be expected from any process or tool that enables decision-making about the design of a building earlier in the project life cycle, such as BIM (Building Information

Modeling) and immersive VR systems do (Dunston et al., 2011). Immersive visualization systems can, for instance, contribute to addressing issues of constructability (Boton, 2018), operability and maintainability of facilities at early design toward minimizing construction, operation and maintenance costs associated with flaws arising from poor solutions (Al-Hammad et al., 1997; Becerik-Gerber et al., 2012). Besides, the adoption of design representations that are closer to the existential-spatial human experience in the real world would contribute to the development of buildings that match end-users demands more effectively, whether technical, functional, or symbolic ones (Florio, 2011).

The assumption that IVR systems provide more effective design representations leads to the expectation that they would also facilitate design review and contribute to problem solving. While the format of representation defines the ease of information understanding and sharing (Chandrasegaran et al., 2013), and the usefulness of a representation depends on how suitable it is for its purposes (Marr, 1982), many scholars argue that the traditionally most used architectural representation formats – two-dimensional (2D) sketches and technical drawings – are similarly limited in terms of their usefulness and ability to convey architecture (Ibrahim and Rahimian, 2010). Two-dimensional drawings require additional cognitive effort in the visualization of the object represented, especially for complex structures (Khemlani et al., 1997). This additional effort is precisely the need for extrapolation of the drawing's scale to one's internal scale, which may require significant training. Two-dimensional drawings may not deliver a good sense of size and volume of space, restricting the ability of stakeholders to understand and suggest necessary alterations to a design (Henry and Furness, 1993). The spatial relationships of an architectural artifact are likely to reach a level of complexity that their representation will benefit from being extended across a third dimension, "allowing the viewer to transverse the data set structure" (Wann and Mon-Williams, 1996).

Another widely adopted method to represent architecture is VR simulations, usually developed with BIM software applications which encompass 3D models displayed through conventional computer screens. VR has become a popular technology for design review due to perceived benefits associated to the representation of scale, depth, and volume (Castronovo et al., 2017). Alternatively, BIM models can be presented within Immersive Environments (IE) – a synonym for IVR technology – where the user can literally walk through and interact with the virtual content (Cruz-Neira et al., 1993). A BIM-enabled IVR system could allow for early development of more integrated design solutions by enhancing the processes of communication, review and evaluation of the implications of solutions developed over the design process.

VR systems can be of different types. Usually, they comprise four main components: a) a computer-generated 3D model (the simulation), b) a display, c) interaction/navigation devices to interact with the virtual model (controllers, data gloves, keyboard, etc.), and d) a software application that orchestrates all the different components. There are different types of display as well, either monoscopic or stereoscopic ones, such as: mobile displays (smartphones and tablets), computer monitors, head-mounted displays (HMD), binocular omni-orientation monitors (BOOM), projection-based panoramic displays, cave automatic virtual environments (CAVE™), 3D glasses (coupled with computer monitors, projection-based, and CAVE-like displays), and virtual retinal displays (VRD).

VR platforms can be categorized into non-immersive and immersive (IVR). Non-immersive, low-end VR systems display monoscopic perspective views of a digital model. Interaction is usually limited to navigation through the environment using a mouse and/or keyboard, which allow movement forward, backward, left and right. In high-end IVR systems, tracking devices detect a user's head and body movements and the display device projects stereoscopic images of a digital model (Bertol, 1997).

The difference between low- and high-end systems resides in the use of interaction and visualization devices that make perception and interaction more natural and realistic according to human perception parameters (Bertol, 1997). Other scholars argue that the most evident difference between non-immersive and immersive VR is precisely the level of presence provided, which the immersive type aims to enhance by employing stereoscopic visualization and intuitive interaction (Steuer, 1992). Since perception runs on input from different sensory channels (Gifford, 2002), it is expected that the more interfaces of interaction with the computer-generated world, the more accurately a VR system would simulate perception. For instance, CAVE-like systems allow users to explore the virtual world using head movements, walking navigation and manipulation of virtual objects with hand gestures (Cruz-Neira et al., 1993). In immersive environments, stereopsis, wide field of view, and high interactivity are all critical elements to generate great levels of presence (Castronovo et al., 2013; Dunston et al., 2011). In sum, depth realism and presence vary largely across low- and high-end VR systems (Bertol, 1997).

Due to the high costs of CAVE-like systems, a few years after the release of the first CAVE™ (Cruz-Neira et al., 1993), low-cost immersive environments began to be developed. These systems usually employ off-the-shelf low-end equipment and require less advanced computational skills to be operated so that they are accessible to a wider range of users (Kalisperis et al., 2002). More recently, the increase in computer processing power over the past two decades allowed for the development and release of a variety of commercial head-mounted displays. A broad range of potential users such as game enthusiasts, researchers, and designers have now access to devices such as the Oculus Rift™ and the HTC Vive™. These platforms usually require very little computational and programming skills and are able to deliver powerful VR experiences. Figure 1 below shows examples of different VR systems.

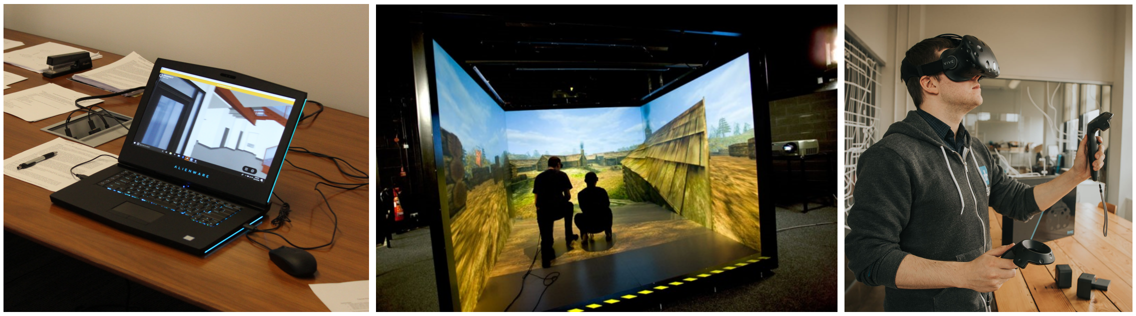


Figure 1. Examples of VR systems: low-end non-immersive (left), high-end immersive CAVE-like (center), and high-end immersive HMD-based (right). Retrieved from www.christiedigital.com (center), www.digitaltrends.com (right).

Regardless of the limited adoption of IVR technology within industry segments (Kalisperis et al., 2002) research on the topic has increased over the past two decades (Renner et al., 2013; Portman et al., 2015). However, when it comes to VR-related research in architecture and construction, the number of studies is drastically lower than in other fields. Comparative, quantitative, user-centered studies are even scarcer. Between 2005 and 2011, out of 150 publications on VR applications in the built environment in leading international journals, an average of 4.6 articles on quantitative effectiveness assessments were published per year. In regards to comparative studies (studies comparing a VR system to a different media), only 2.7 publications were published per year on average (Kim et al., 2013). Due to the lack of quantitative and/or comparative studies that clearly indicate improvements to user performance there is still much controversy about the effectiveness of IVR systems in various construction applications. The main reason for the lack of studies in this topic may be the fact that for many potential researchers VR continues to represent a complicated subject, requiring them to work far beyond the boundaries of their original knowledge fields (Otto et al., 2003).

Some previous studies on IVR applications in construction – mostly qualitative, exploratory, and case studies – explored the benefits that the technology can offer to the design process. Traditionally, in the development of designs of specialized buildings, designers would build physical full-scale mock-ups to collect end-user feedback on different issues such as visual sightlines ("vistas") and access to key equipment (Majumdar et al., 2006; Kumar et al., 2011). Alternatively, one could use immersive digital mock-ups. Past studies suggest that IVR systems are effective simulation tools (when virtual experiences are compared to real-world experiences), and therefore have a strong potential to be used in research and architectural design (Kuliga et al., 2015). IVR-enabled experience-based design review has proved successful in supporting decision-making in the delivery of complex projects (Leicht et al., 2010). In IVR-based design review meetings conducted by Dunston et al. (2011), most modifications proposed to the original design of a healthcare facility related to the spatial arrangement of equipment and workstations, providing evidence that IVR systems may benefit the understanding of spatial relationships of a 3D model. Dunston et al. (2011) also noticed that when using an IVR system, stakeholders from different disciplines were able to provide early and detailed contributions. Maldovan and Messner (2006) showed that an immersive system can reduce the time of descriptive and explanatory conversations and increase the time of evaluative and predictive conversations (compared to traditional project meetings, without the support of immersive visualization), enabling the generation of a greater number of alternative design solutions and increasing the likelihood of a better final design.

Overall, evidence suggests that immersive environments provide the conditions for better collaboration among stakeholders (Majumdar et al., 2006; Bassanimo et al., 2010; Kumar et al., 2011; Fernando et al., 2013; Berg and Vance, 2016), enable better understanding of prototypes (Berg and Vance, 2016) in comparison to 2D media and non-immersive systems (Schnabel and Kvan, 2003), allow for the anticipation of design decisions and the identification of design issues that would not be identified otherwise (Okeil, 2010; Bassanimo et al., 2010; Dunston et al., 2011), and can also predict human-building interactions that can feed designers and researchers with reliable user behavior data (Bassanimo et al., 2010; Adi and Roberts, 2014; Kuliga et al., 2015; Heydarian et al., 2014, 2015a, 2015b, 2015c). Thus, costs associated with time spent on decision-making and with the manufacturing of physical mock-ups used in design review could be significantly reduced with the use of immersive systems (Majumdar et al., 2006; Maldovan and Messner, 2006). Safety training also benefit from the adoption of such systems (Xie et al., 2006; Sacks et al., 2013), probably due to their ability to promote high levels of involvement, which contribute to the learning process (Faas et al., 2014). Wang et al. (2013) stated that integrating facility managers into early design stage with support of advanced 3D visualization could benefit collaborative space planning and significantly reduce a building's life cycle cost by minimizing the chances of major repairs that would otherwise occur at the operational phase. Also, VR can better inform and support decisions made by facility managers when performing maintenance and repairing tasks by providing ubiquitous intuitive 3D visualization of an as-built BIM model through mobile immersive VR platforms (Yang and Ergan, 2015), representing a "powerful means to retrieve information from a virtual model of a facility" (Gao and Pishdad-Bozorgi, 2018).

To some extent, all these perceived benefits refer to the aforementioned effectiveness of building representations provided by IVR technology. In the context of design review, such effectiveness seems associated with realism of immersive simulations. Immersive visualization is expected to provide users with "a realistic perception of the design" (Fernando et al., 2013) and "simulate the experience of moving through and interacting with the virtual world as if it was real" (Bassanimo et al., 2010). As per Bertol (1997), the main goal of a VR simulation in design applications is to resemble a future, envisioned architectural environment.

Realism is one of the major achievements of computer-generated models. However, the VR equipment consisting of high-end software and displays does not create a realistic VR experience; instead, it allows people to construct the scene by themselves, similarly to when looking around a physical environment. The human visual intelligence is in charge of creating and qualifying visual experiences in the computer-generated world (Hoffman, 1998). It follows that perceptual criteria dictate the effectiveness of virtual environments (Wann and Mon-

Williams, 1996). Then, what makes a VR experience more compelling or "realistic" might be its ability to allow users to form more realistic mental images, that is, "percepts" (or yet, "constructs") that are to a certain extent more similar to the ones they would create when at the physical world. Because human experiences are a function of one's perceptual processes (Gibson, 1966; Hoffman, 1998), the effectiveness of a VR system for design review purposes can only be established in terms of the resemblance between human perceptions in the virtual environment *versus* human perceptions in the correspondent physical environment, that is, the extent to which the virtual environment mimics one's perceptual processes in the physical reality. Ultimately, the advantages of using IVR technology to represent the spatial features of architectural objects are only relevant if spatial perception of virtual environments is similar enough to spatial perception of real-world spaces. Only by satisfying this condition it is possible to draw conclusions regarding the effectiveness and benefits of IVR technology with respect to architectural representation.

In the visual perception of a fundamentally three-dimensional world, depth perception plays a critical role. Thus, the level of realism of depictive representations is largely affected by the extent to which these are able to convey depth. Historically, the representation of depth in depictions of architecture has been intensively explored and pursued through a variety of techniques such as perspective constructions, in an effort to provide observers with realistic architectural representations. Depth representation and perception are expected to enhance realism of architectural representations and facilitate the understanding of the spatial relationships of envisioned architectural artifacts. Realistic representations were and continue to be necessary to translate architectural compositions into perceivable artifacts and communicate ideated spaces (Bertol, 1997). The ability to visualize shape, size, volume, proportions and scale of spaces is prerequisite for solving spatial problems in the architectural design process (Zikic, 2007). Understanding the spatial characteristics of a future architectural artifact from its representation is critical in the design review phase, as it deals with the communication, review and evaluation of the implications of spatial solutions to a project's technical, functional, and symbolic demands (Dunston et al., 2011; Florio, 2011). These relationships are illustrated in Figure 2 below. In this context, immersive environments are expected to provide scale and depth perceptions comparable to perceptions of the real world (Zikic, 2007).

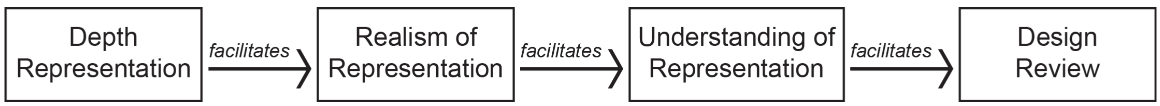


Figure 2. Relationship between depth representation and design review

It is important to highlight that in this work IVR technology is not deemed a tool to create design solutions, that is, a tool for the creative, problem-solving act of design, which is still up to a designer's creativity, intelligence, experience, and background – although it may be up to computer algorithms eventually. There is a clear-cut distinction between design and representational tools. An IVR system does not solve problems by itself, but allows for the communication, review and evaluation of the implications of solutions developed – a collaborative process that usually involves the participation of several professionals and clients, not only designers. In that sense, IVR is deemed a representational tool for communication of design solutions in the design review process and not for the creative process of ideation, in which designers make use of diverse techniques such as schematic diagrams, hand sketches, and physical prototypes.

The unquestionable efficiency of electronic drafting and three-dimensional modeling brought by CAD technology does not give a legitimate reason to state that computers actually "aid" in the ideation process, which makes the expression "Computer-Aided Design" inadequate (Bertol, 1997). Similarly, VR systems do not investigate design alternatives by themselves, but can be used in design review to create interactive walkthroughs, providing a building simulation understandable by construction agents involved in the act of discussion and evaluation of the solutions proposed. Walkthrough navigation is the primary means of exploration of architectural simulations and can reveal design issues that would not be detected otherwise (Bertol, 1997).

Presence is also expected to benefit the design review process, as it has been shown to correlate significantly with information acquisition, learning, and involvement in virtual environments (Oren et al., 2012; Faas et al., 2014). Acquisition of visual information in virtual environments is essentially a visual perception process, which involves directed attention orienting one's sight toward visual information sources and selective processing of available information (Gibson, 1979; Gifford, 2002). In the context of design review, greater levels of presence are expected to provide users with enhanced ability to search, locate and identify visual information (also known as visual search), improving visual perception of their virtual surroundings (Kalisperis et al., 2006; Heydarian et al., 2015). As discussed throughout this dissertation, presence measurement methods can be controversial as it results from a complex relationship between VR system properties and human factors. Identifying and characterizing technological and human factors that affect presence has been acknowledged by several scholars as a critical step towards the development of VR systems to enhance human capabilities in various contexts (Slater, 1999; Zikic, 2007; Interrante et al., 2008).

Current approaches to address the effectiveness of VR systems for checking the conformity of design solutions are not on the basis of effectiveness of solutions per se (it is rather difficult to determine quality of design solutions), but either on the effectiveness of the design process (task-centered approaches utilizing time and cost metrics) or on the characteristics of cognitive processes involved in the design review task (user-centered approaches utilizing cognitive performance metrics) (Kim et al., 2013).

In order to determine the value of IVR technology in construction applications, a comprehensive study could employ either or both aforementioned approaches. However, there is an evident and concerning gap between VR-related research in the fields of cognitive sciences and construction. This gap refers to the consideration (or lack thereof) of visual perception, sense of presence, and other critical cognitive processes inherent to VR experiences. Current studies on IVR applications in construction are mostly qualitative, exploratory or case studies, aiming at exploring implications on productivity, task performance time and other effectiveness indicators (Kim et al., 2013), neglecting the cognitive benefits that the technology may provide – an issue that comes before productivity – and disregarding human factors inherent to discussions on VR applications (Paes and Irizarry, 2016). As per Higuera-Trujillo et al. (2017), research in the field is limited in three major aspects: a) obsolescence of the studied VR platform, which stresses the importance of critically and comparatively updating their validity, b) disregard of cognitive aspects of user experience that underlie human behavior and psychological state in virtual environments, and c) lack of adoption of user's objective responses in validation studies.

In other words, many studies on virtual modeling and advanced visualization overlook the cognitive processes that underlie the understanding of computer-generated 3D representations, focusing on the analysis and description of systems' properties instead. In contrast, there is a growing number of empirical studies in the cognitive sciences on human and technological factors that may affect perception and presence in virtual environments (Renner et al., 2013). The most widely accepted definition of VR – one's experience of presence in virtual worlds (Steuer, 1992) – places the human user as its central, defining component (Stanney et al., 1998). Thus, any analysis of effectiveness of virtual environments should be conducted on measures of user experience/performance, in conformity with main references in the fields of human-computer interaction (HCI) and cognitive psychology.

A fundamental step for IVR to be considered an effective technology for collaborative design review is to understand whether and to what extent it enhances users' cognitive abilities, that is, their performance in obtaining and understanding the spatial relationships depicted, or yet, their three-dimensional perception and sense of presence (Wann and Mon-Williams, 1996; Stanney et al., 1998). As per Cutting and Vishton (1995), our perceptual judgments are a function of both perceived environment and perceiving subjects. Thus, only by understanding whether, to what extent, in what circumstances, and to whom virtual environments are beneficial it will be possible to define the ergonomic, environmental, technological and representational parameters upon which such systems would be efficient at mimicking the physical world and, ultimately, benefit the design process. As suggested by Wann and Mon-Williams (1996) and Chandrasegaran et al. (2013), virtual environments should be designed around human perceptual capabilities in the context of the task to be performed. A systematic investigation on human factors that influence perception and presence in virtual environments could provide developers with reliable information to design systems tailored for specific applications or able to adapt to a broader range of users (Stanney et al., 1998; Okamoto, 1999).

CHAPTER 2

MOTIVATIONS & RELEVANCE

The general motivations of this research involve testing the assumptions regarding the effectiveness of IVR technology in the design process. It aims to produce and bring evidence into the general debate on the effectiveness of immersive systems for collaborative design review by investigating the existence and extent of improvements in 3D perception and presence levels of a specific user population. Past research on virtual modeling and advanced visualization in the AECO-FM domain has not demonstrated the benefits of IVR systems to architectural design in comparison to traditional non-immersive media. Systematic comparisons to non-immersive VR systems are critical to support or refute the assumptions concerning the benefits of the adoption of immersive environments in collaborative design review, as well as to investigate the validity of using IVR technology in lieu of non-immersive systems. The lack of knowledge on that matter represents a threat to work environments that use IVR systems for collaborative design review. Also, the assumption that immersive systems would reproduce physical world conditions accurately, under all circumstances, for all purposes and for everyone, may not contribute to expand knowledge on the true usefulness and contributions of immersive visualization to the AECO-FM industry.

The relevance of this research lies in the generation of knowledge on the extent to which IVR technology offers better support to design review compared to non-immersive VR platforms based on improvements to 3D perception and presence. This knowledge is essential to elicit the advantages and effectiveness of immersive representations and to provide directions to the development and implementation of increasingly effective IVR systems across the industry. Exploratory and qualitative studies were crucial to lay down the foundations of research in the field. However, because these were mostly inspection-based, task-centered, and self-evaluation studies, they were not able to provide much evidence on whether and to what extent IVR systems are able to deliver more effective representations than traditional representational media. Therefore, many studies have been conducted on a highly contentious assumption that immersive environments are inherently more effective in conveying geometric information despite the lack of evidence to support such statement, and would often deliver a rather anecdotal report of users' experiences with IVR systems. Despite numerous user-centered studies from other knowledge domains (e.g., human-computer interaction, cognitive psychology, and computer science) on the extent to which immersive simulations are similar to physical environments, there is still not enough evidence to support the hypothesis that immersive environments could replicate one's perception of a real environment effectively. In addition, and in terms of the relevance for the design practice, very few studies discuss whether that accuracy is higher than using traditional non-immersive visualization media. The combination of four methodological characteristics in this study (context, comparative, quantitative, and user-centered approach) is expected to address these issues. Experimental research appears to be the only approach that enables researchers to make judgments about beliefs and assumptions with systematically measured confidence and reliability (Lazar et al., 2017).

CHAPTER 3

OBJECTIVES & QUESTIONS

The general research hypothesis is that an IVR system would be able to enhance users' three-dimensional perception of a BIM model and sense of presence in the virtual environment in comparison to a conventional non-immersive VR system. A non-immersive VR system (a BIM model displayed through a laptop screen) is selected as the condition for comparison since it has been the most used VR system in the industry for the past two decades (Bertol, 1997).

The research questions are as follows: 1) Does the IVR system enhance users' three-dimensional perception of a BIM model? 2) Does the IVR system enhance users' sense of presence in a BIM-based virtual environment? 3) Is there an association between presence and three-dimensional perception in virtual environments? The statistical hypotheses H1, H2 and H3 (listed in Table 1) aim to answer research questions 1, 2 and 3, respectively.

Table 1. Statistical hypotheses

Hypothesis	Variables		Null hypothesis H_0	Alternative hypothesis H_1
	Independent	Dependent		
H1	VR modes: niVR and IVR	3D perception	There is no difference in 3D perception (dep.) between VR modes (ind.)	There is a difference in 3D perception (dep.) between VR modes (ind.)
H2		Presence	There is no difference in presence (dep.) between VR modes (ind.)	There is a difference in presence (dep.) between VR modes (ind.)
H3		3D perception and Presence	There is no relationship between 3D perception and presence (dep.) in the VR modes (ind.)	There is a relationship between 3D perception and presence (dep.) in the VR modes (ind.)

In statistical terms, the research goal consists in finding out statistical evidence to refute or nullify the null hypotheses in order to support the alternative hypotheses. More specifically, the experiment aims at finding out whether and how changes in a single independent variable (VR mode) of two different values, namely, the niVR (a BIM model displayed through a laptop screen) and the IVR (an immersive virtual reality system) modes, induce changes in dependent variables, namely, 3D perception and presence, controlling for user-related confounding variables, that is, individual characteristics of age, gender, educational level, bachelor's major, current major, experience in design review, computer usage, experience with 3D virtual environments, familiarity with the experiment environment, and spatial ability. Measures of 3D perception and presence in each experimental condition are used to perform a direct comparison between niVR and IVR systems. The dependent and independent research variables are described in Table 2 below.

Table 2. Research variables

Variable type	Name	Values	Description
Independent	1. VR mode	niVR	A BIM model presented through a conventional laptop screen, i.e., a non-immersive VR system.
		IVR	A BIM model presented through a commercial head-mounted display, i.e., an immersive VR system.
Dependent	1. 3D perception		The extent to which a participant has an accurate 3D perception of the environment depicted. The accuracy of 3D perception is given by the ratio of hits over the total number of observations. A hit is defined as when a participant chooses the same answer to a given question on the 3D perception questionnaire in both reference and VR mode. In other words, accuracy refers to the deviation of a participant's 3D perception in VR modes against her/his perception in the real world.
	2. Presence		A participant's level of involvement and immersion in a virtual environment.
Confounding	1. Age		Individual factors believed to affect dependent variables. These are controlled by adopting a within-subject design (Thompson & Campbell, 2004), as well as through statistical analysis (Pourhoseingholi et al., 2012).
	2. Gender		
	3. Educational level		
	4. Bachelor's degree major		
	5. Current academic major		
	6. Experience in design review		
	7. Computer usage		
	8. Experience with 3D virtual environments		
	9. Familiarity with the experiment environment		
	10. Spatial ability		

3.1 Research Question # 1

Does the IVR system enhance users' three-dimensional perception of a BIM model? This question is associated with statistical hypothesis H1 and investigates the relationship between virtual reality systems and three-dimensional perception of a BIM-based architectural model.

In the design of specialized facilities and increasingly complex buildings, designers and end-users are oftentimes required to assess if dimensions and proportions of a space are adequate for its purposes – whether it is a hospital surgery room, a religious temple, or a circulation area. This assessment could be improved in a scenario where those agents are able to walk through a virtual building and test the adequacy of its dimensions and proportions in light of various demands (ergonomic, technical, functional, symbolic, etc.), and suggest necessary alterations (Florio, 2011; Dunston et al., 2011).

During design review, professionals and clients are usually concerned with the implications and efficiency of dimensions and proportions proposed, regardless of their numerical measure. Initially, stakeholders need to evaluate whether these dimensions, proportions, shapes, and areas would work or not – an oftentimes subjective assessment. If the focus is on designing spaces for optimal use, VR simulations can aid in the determination of layout configurations, room dimensions and proportions, sizes of walls and openings, and height and slope of ceilings in regards to the activities to be undertaken in the future building (Schnabel and Kvan, 2003; Dunston et al., 2011; Kuliga et al., 2015; Berg and Vance, 2016). End-users can check for the ergonomics and conformity of solutions providing feedback on the adequacy and functionality of equipment location, circulation areas, etc. (Bertol, 1997). A client may want to "feel" how spacious a sitting room is; a landscape architect may check if there are any visual obstructions to a landmark of interest from a certain vantage point; nurses and surgeons may verify and provide feedback on the accessibility to key equipment in a surgery room; a facility manager may want to check whether there would be enough room to perform a certain repairing procedure on the

HVAC machinery (these examples are provided by Al-Hammad et al., 1997; Maldovan and Messner, 2006; Majumdar et al., 2006; Leicht et al., 2010; Kumar et al., 2011; Wang et al., 2013; Yang and Ergan, 2015).

In immersive VR simulations, dynamic perceptions through active walkthrough exploration and changing of the viewpoint (usually kept at human height) and direction of sight can convey the feeling of inhabiting a room, the sense of closeness or openness when walking down a street, and the visual appearance of proportions of architectural elements. VR walkthroughs provide a simulation of what people would visually perceive walking through a built space, and can be useful to visually test the efficiency of design alternatives at the human scale (Bertol, 1997).

Often taken for granted, that feeling of scale of a space is in reality sustained by unconscious and dynamic processes of spatial perception. Although the relevance and purposes of spatial perception in design review may vary across construction professionals and their specific information needs, the importance of shared understanding of a project's design by all stakeholders involved in the design phase – not only experienced designers – is unquestionable for achieving more integrated solutions. Understanding the spatial relationships of an architectural model is critical in the collaborative design review task (Dunston et al., 2011; Fernando et al., 2013). Only by understanding the architectural representation, designers, construction agents, clients, and end-users would be able to adequately evaluate a design and provide feedback on the implications of solutions developed.

Realistic representations continue to be necessary to translate architectural compositions into perceivable artifacts and communicate ideated spaces (Bertol, 1997). These are expected to benefit the understanding or architectural designs and facilitate design review (Figure 2, Chapter 1). The level of realism of architectural representations is largely affected by the extent to which these are able to convey depth. The expectation is that depth perception would be enhanced by stereoscopic visualization enabled by IVR technology, meaning more realistic representations.

In essence, IVR technology aims to allow people to visually perceive a virtual world as they perceive the physical reality (Bertol, 1997). Interestingly, the higher realism of IVR-based architectural representations in terms of depth perception is still an assumption and has not been tested to date. This level of realism of virtual representations is given by the accuracy of depth perception with respect to perception in the real world. Accurate depth perception approximates the architectural representation to the future architectural artifact being represented.

Naturally, depth representation per se does not guarantee accurate perception of the three-dimensional structure of a space. One must perceive depth from its representation. A given architectural representation can only be deemed more realistic if one actually perceives three-dimensionality more accurately from it. Therefore, in reality, there is something in between depth representation and realism or architectural representations shown in Figure 2 (Chapter 1) and that is precisely three-dimensional perception (Figure 3), the dependent variable in hypothesis H1.

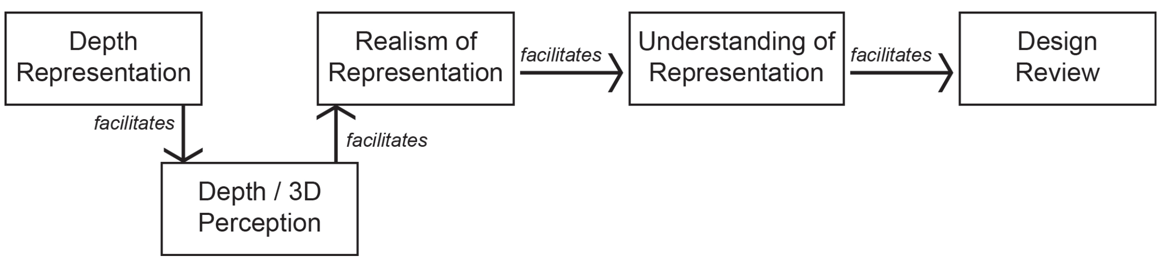


Figure 3. Relationship between depth representation and realism of representation

Due to their direct association, distance estimation has been widely used as a metrics of depth perception in virtual environments (e.g., Witmer and Sadowski, 1998; Sinai et al., 1999; Gooch and Willemsen, 2002; Thompson et al., 2004; Interrante et al., 2006; Renner et al., 2013). Therefore, this question aims at verifying if an IVR system would allow users to estimate distances more accurately, i.e., to have more accurate three-dimensional perception of an architectural design, in comparison to non-immersive VR technology. In other words, three-dimensional perception is operationalized in this study as one’s understanding of the dimensions, proportions, and scale of an architectural design (Zikic, 2007), provided by estimates of egocentric distances to objects in space and distances between objects in space (horizontal, vertical, and depth judgments combined). In order to avoid any misconceptions of nomenclature, this study adopts the term "three-dimensional perception" or simply "3D perception", as did Wann and Mon-Williams (1996) and Norcia and Gerhard (2015), to refer to a particular perceptual process within visual perception that governs the interpretation of visual sources of information about the three-dimensional structure of a space, that is, its three-dimensionality (X/horizontal, Y/vertical, and Z/depth dimensions). In sum, this question compares the ability of two distinct VR systems (non-immersive *vs.* immersive) in conveying the three-dimensionality of the architectural space represented.

3.1.1 Visual Perception

Perception is defined as the result from the interpretation of external stimuli captured by our senses. It is an intricate self- and spatial awareness phenomenon that involves directed attention orienting one's senses toward information sources and selective processing of information available (Gibson, 1979; Gifford, 2002). As per Gibson (1966), the process of perception is a function of inputs from various sensory channels, as well as of more mindful mental processes that assimilate incoming sensory data with current concerns and past experiences.

Visual perception is amongst the various types of perception. Theories of visual perception are numerous; some of which date back to the ancient Greece. Hoffman (1998) claims that people do not see the reality as it is; instead, they construct what they see, including colors, shapes, and scenes. When visually perceiving their surroundings or even an external depictive representation (the type of representation that conveys meaning through its resemblance to the object it aims to represent), people produce a mental representation of it, oftentimes referred to as "constructs" or "percepts" (Kosslyn, 1994; Hoffman, 1998; McGinn, 2004). These constructs would be triggered by external visual stimuli. People would "unconsciously" infer visual properties such as distances, sizes and shapes of visible objects in space, although the fabrication of constructs/percepts – the mental representations resulting from visual perception – would happen in stages, obeying certain rules named "innate rules of universal vision." Moreover, although vision could be deceived in the fabrication of constructs by deliberate illusory depictions, it cannot be deceived so easily in real life (Cutting and Vishton, 1995). Figure 4 shows a schematic diagram showing the process of perception of VR representations.

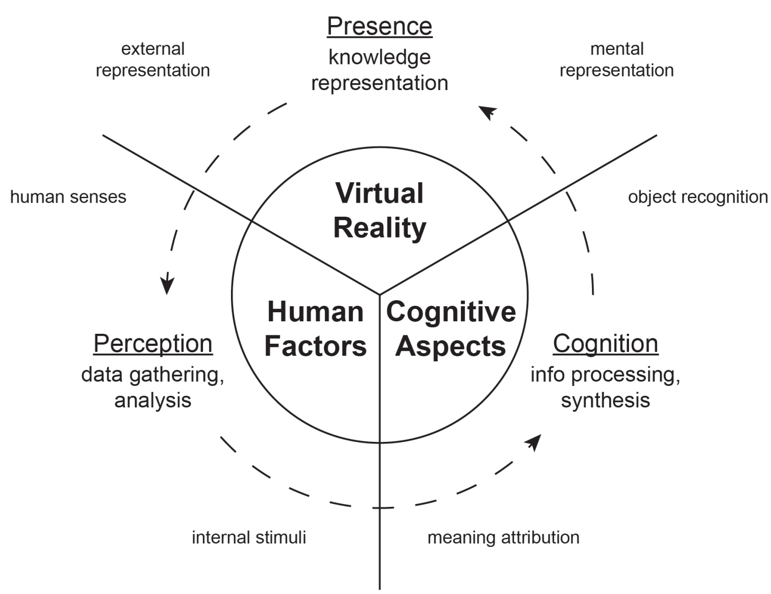


Figure 4. Perception process of VR representations. Adapted from Paes and Irizarry, 2016.

Mental imagery will have different properties depending on the type of depictive representation in use (Kosslyn, 1994). Regardless, people have evolved to construct objects and spaces in three-dimensions (Hoffman, 1998), according to the laws of physics, and as they would exist physically. Although our imagination may not be limited to the behavior of things in the physical world, Hochberg (1998) states: "*a subject's mental representation is isomorphic to the physics of real objects*." Naturally, traditional representation methods cannot thoroughly convey all information embedded in one's mental image. For instance, static perspective renderings cannot afford certain visual cues or sources of information deemed critical in the perception of three-dimensional artifacts and spaces such as motion perspective, motion parallax, and stereopsis. This argument supports the expectation that the adoption of representation and visualization methods that are able to convey more accurately the results of "visual" perception of mental images would benefit design, as the most relevant aspect of an efficient design representation tool seems to be its ability to faithfully correspond to, reproduce, and communicate a designer's idea. As Gibson (1979) stated, "*two-dimensional images are rich in information, but moving images are richer still*."

Visual perception is a sophisticated, creative and innate process, through which people tend to construct three-dimensional objects and worlds. Evolutionarily speaking, the human being is trained to see things in three-dimensions. When looking at a static figure of a house in perspective, for instance, the human brain will likely understand the three-dimensional shape of the house from that flat representation (Marr, 1982; Hoffman and Singh, 2006). In the process of constructing three-dimensional scenes and objects, the visual system makes use of multiple "sources of information" or "visual cues", including: two-dimensional contours, shading, texture, occlusions, stereovision, binocular disparity, motion parallax, motion perspective (when the observer is moving, objects move faster if they are closer, slower if they are farthest), aerial perspective (the farthest objects appear bluish), height in the visual field, relative size, relative density, and prior knowledge (Cutting and Vishton, 1995; Hochberg, 1998).

In everyday tasks, people have to successfully interpret a multitude of visual cues, which provide information about size, shape, color, location, and other aspects of objects and spaces (Zikic, 2007). As per Cutting and Vishton (1995), the more diverse information sources are, the more accurate and consistent visual perception. The relative importance of visual cues, however, is still unknown, but they certainly interact, reinforce, conflict with each other and build on one another. Environments are typically rich in sources of information, but they can also be extremely varied, with certain kinds of information present in some situations but not in others. For instance, at different distances, some sources fail to deliver an adequate quality of information (some sources vary or fade out with distance) so that the visual system is forced to rely on others (and one may still fail to perceive).

A particular visual process within visual perception namely, depth perception, accounts for the perception of depth of objects in space. Visual spatial cues can be broadly categorized as either geometric (e.g., distance, direction) or featural (e.g., color, texture) (Kimura et al., 2017). Depth cues can be either geometric or featural, and are further categorized into primary cues (stereopsis and parallax) and secondary cues (motion parallax, linear perspective, occlusion, size, texture, shading and shadow, light, color, among others) (Kelsey, 1993; Khuu et al., 2014). Provided by binocular vision, the binocular depth cues of stereopsis, parallax, and binocular disparity (the difference in the positions of binocularly visible objects) are deemed by many scholars as the most relevant depth cues, although their relative importance is still unclear (England et al., 1992; Hubona et al., 1997; Hubona, 1999; Brooks, 2017). Wann and Mon-Williams (1996) provide an investigation of interactions among depth cues. It appears that in the absence of primary depth cues, depth perception relies on secondary depth cues. Contrary to the expectations of many VR scholars, stereopsis may not be so important to make judgments about depth in the real world – occlusion and perspective, which are provided by monocular vision, may be even more important (Cutting and Vishton, 1995).

Although the processes of visual perception of two-dimensional images are relatively well known (e.g., Marr, 1982; Kosslyn, 1994; Palmer, 2003), the ones involved in the visual perception of dynamic three-dimensional environments – such as of real-world and VR experiences – are still unclear (Stanney et al., 1998). Contemporary scholars may refer to this particular process as perception of layout or "layout perception", which encompasses the idea of perceiving the arrangement and displacement of objects in space. The general argument is that people do not perceive space but objects in space, invalidating the expression "spatial perception"

(Cutting and Vishton, 1995; Gibson 1979), although adopted by many scholars (e.g., Henry and Furness, 1993). Alternatively, researchers have utilized a variety of terms to refer to the process of perception of the spatial configuration and three-dimensional arrangement of environments such as "spatial understanding" (Schnabel and Kvan, 2003), "spatial cognition" (Kimura et al., 2017), "3D visualization" (Hubona et al., 1997), or yet, "spatial comprehension" (Zikic, 2007).

3.1.2 Perception of Three-dimensionality from Architectural Representations

Daniela Bertol (1997) provides a thorough analysis on this subject. She starts by stating that the medium of architecture is fundamentally three-dimensional space. Works of architecture are often three-dimensional solid artifacts generated by the shaping of space defined by boundaries. In the most basic architectural artifacts, boundaries consist of elements such as floor, ceiling, and walls. The identification of boundaries, of being inside or outside, and of the entire architectural artifact is subject to the perception of the space created by architecture. The description of such relationships in architectural representations for the purposes of visualization and communication of envisioned spaces becomes challenging due to the limitation of two-dimensional representation techniques. However, the representation of the three-dimensionality of architecture through 2D media was improved with the advent of perspective constructions, which aimed at "better representing depth" (Bertol, 1997).

The advent of perspective was prompt by the quest for realistic representations, which were necessary to translate architectural compositions into perceivable artifacts and communicate ideated spaces. It connected the bridge between representation and reality to the bridge between reality and ideas, becoming one of the most powerful means of design: ideas could now be pictorially communicated in an accurate and understandable fashion. The expression of depth through perspective provided more realistic visual reproductions of reality, and improved expression of ideated environments. Since the Renaissance period, perspective constructions have been used to create the illusion of the third dimension in architectural representations on two-dimensional media, providing the most accurate visual simulations of how an envisioned architectural artifact would appear to the eyes in the real world.

Whether a depictive representation is able to describe and convey the location of points in space in a more accurate way – accuracy as being the degree of resemblance to location of points in the physical world – the mental representation created upon its perception will be closer to the physical reality, space, or object it aimed to describe. Thus, the question resides on whether our visual perceptions of the real world are similar to our perceptions of a virtual world mapped from it (Figure 5). In other words, are our perceptions of a virtual environment similar to our perceptions of a physical environment? The accuracy of this mapping process gives the compatibility between perceptions of physical and virtual realities, and establishes the visual realism of virtual environments.

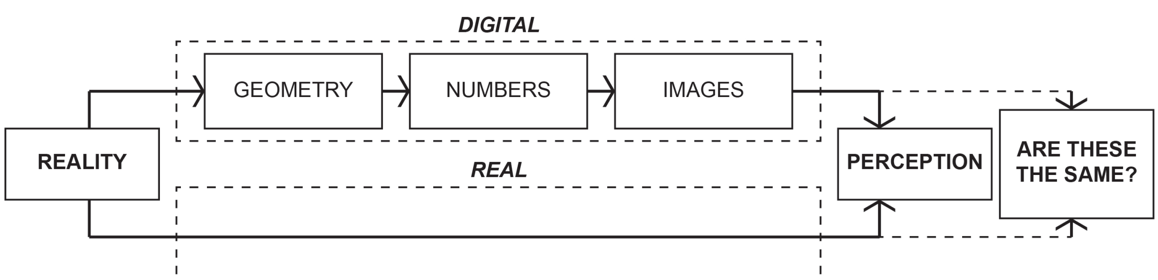


Figure 5. Compatibility between perceptions of virtual and physical worlds. Adapted from Bertol, 1997.

3.1.3 Perception of Three-dimensionality from VR-based Architectural Representations

Perspective constructions represented a major achievement in the representation of depth and increment of realism of architectural representations. However, in order to achieve a complete illusory effect from static perspective renderings the observer needs to be standing in a fixed position coincident with the viewpoint from which the perspective was developed.

Monocular observation was also a requirement to compensate for the perceived lack of depth, informed by an observer's binocular vision. The point is that perception of built architecture is not static, neither monoscopic, but stereoscopic and dynamic: the optimal contemplation of an architectural artifact is provided by the change of perspectives over time and space, providing a sequence of views. What once was the goal of perspective constructions – to create the illusion of reality – was recently leveraged by VR technology, which is able to deliver representations that are simultaneously three-dimensional and dynamic (Bertol, 1997). One of the earliest definitions of contemporary virtual environments described them as "interactive 3D computer animations structured within an in-depth three-dimensional space" (Wann and Mon-Williams, 1996).

When a VR model of an architectural design is displayed and explored through a conventional computer monitor, the visual system interprets the images as a sequence of two-dimensional perspective renderings. There is an evident distinction between the situation where an observer perceives her-/himself within and pertaining to a structured environment, and where she or he perceives the visual input as two-dimensional projections of a three-dimensional animated object, such as a digital model of a house viewed through a flat display (Wann and Mon-Williams, 1996). The brain may still monocularly infer depth from 2D animations of 3D objects using certain visual cues such as motion parallax (movement between observer and observed object) and linear perspective (variation of the size of images of the observed object) (Hoffman, 1998). However, other visual cues are needed for an accurate depth perception, such as stereopsis. Stereopsis acts as an important depth cue and further augments spatial visualization (Zikic, 2007). In order to capture depth information through stereopsis one needs to use binocular vision. Binocular vision enables the human brain to reconstruct a three-dimensional world from a pair of two-dimensional retinal images captured simultaneously from two different viewpoints (each eye) (Chen et al., 2012).

Immersive VR display systems recreate stereopsis by providing stereoscopic images to achieve the most realistic visuals. Historically, VR technology has always been associated with the visual phenomenon of stereopsis, as shown by the earliest VR artifacts. Stereoscopic images are one of the main characteristics of VR, differentiating if from other types of computer visualization in the achievement of greater perceptual realism. Current VR systems are but the latest in a series of devices developed to enable the perception of stereoscopic images, from early stereoscopes to modern head-mounted displays (Bertol, 1997).

Naturally, a virtual environment cannot deliver the entire spectrum of visual cues. In the absence of many important ones, visual perception in virtual environments would rely on a lower amount and diversity of visual cues in such a way that the ones that are present would be critical. The most important visual cues to reproduce the three-dimensionality of the world might be the ones that allow for depth perception, such as occlusion, stereopsis and perspective. Consequently, these cues should be included in a virtual environment if its purpose is to provide accurate depth perception. This is particularly true for stereopsis, which is expected to greatly affect depth judgments (England et al., 1992).

The possibility of depth perception through stereoscopic visualization is a defining factor of immersive VR systems and represents a real breakthrough for VR technology (Bertol, 1997). Nonetheless, little is known about whether or not stereoscopic visualization actually benefits perception of depth from virtual architectural representations.

3.2 Research Question # 2

Does the IVR system enhance users' sense of presence in a BIM-based virtual environment? This question is associated with statistical hypothesis H2 and investigates the relationship between virtual reality systems and presence in a BIM-based architectural model.

Architectural artifacts are essentially not only three-dimensional but immersive as well; architecture can be inhabited and walked through on its inside. The inherent characteristics of three-dimensionality and enclosure of architectural artifacts find correspondence in immersive environments, which enable the experience of inhabiting a space before existing physically (Bertol, 1997). The subjective human experience of "being in" a given environment, derived from perceptual processes, is precisely what scholars call sense of presence. Witmer and Singer (1998) define presence as "the subjective experience of being in one place or environment, even when one is physically situated in another" and argue that it is based on "the interaction between sensory stimulation, environmental factors, and internal tendencies." Presence is considered a central aspect of VR experiences, the result of perceiving oneself in the digital space (Wann and Mon-Williams, 1996), as well as an essential component for the complete perceptual experience in the virtual world (Bertol, 1997). Discussions on presence principles and assessment methods were boosted with the increase of studies on virtual reality; nevertheless, little is known about it to date (Kalawsky et al., 1999).

Great levels of presence in virtual environments are expected to benefit design review as presence has been shown to correlate significantly with information acquisition, learning (Oren et al., 2012), task performance, and even quality of design solutions, although one may still perform well while experiencing low levels of presence (Faas et al., 2014). Great levels of presence are expected to provide users with enhanced ability to perform visual search (search, locate and identify visual information), improving their visual perception of virtual surroundings and facilitating the identification of design issues (Kalisperis et al., 2006; Heydarian et al., 2015). Figure 6 illustrates the relationship between presence and design review.

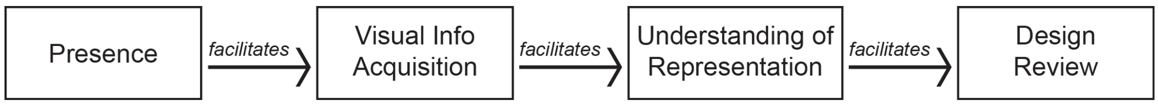


Figure 6. Relationship between presence and design review

Witmer and Singer (1998) argue that the sense of presence in virtual environments is a function of two essential factors, namely, involvement and immersion, which in turn are subject to individual factors (including immersive tendencies) and a virtual environment's characteristics – such as the properties of the visual stimulus (resolution, color, sharpness, brightness, contrast, etc.). Information presented via other sensory channels may also contribute to the experience of presence, but perhaps to a lesser extent than visual information. Studies have consistently found that individual factors can significantly affect presence (Nowak et al., 2008); level of experience and age are known examples (Stanney et al., 1998). Contrary to Witmer and Singer's (1998) expectations, Khashe et al. (2018) found no relationship between the characteristics of VR platforms and presence levels. In Witmer and Singer's studies (1998), however, there is a significant degree of ambiguity in the description of factors thought to underlie presence. These comprise control, sensory, distraction and realism factors, and are expected to interact with one another and may influence presence by affecting involvement, immersion, or both. It should be noted that most of these factors have never been verified empirically.

3.3 Research Question # 3

Is there an association between presence and three-dimensional perception in virtual environments? This question is associated with statistical hypothesis H3 and investigates the association between three-dimensional perception and presence in a BIM-based architectural model. Its goal is to verify whether 3D perception and presence are associated within each VR mode, separately. In other words, to test whether the dependent variables are significantly related in each virtual environment, as suggested in the literature.

People working in the area of presence are trying to unveil the factors that promote it (both human and technological factors, i.e., characteristics of a VR system such as frame rate, pictorial realism, and interactivity) (Khashe et al., 2018). One may find out an equation that allows the trade off among factors towards increasingly immersive VR experiences (Slater, 1999). Despite the relatively large amount of research in the field, there is still much controversy about which factors affect presence in virtual environments (Renner et al., 2013). In sum, identifying and characterizing the technological and human factors that affect presence has been acknowledged by many scholars as a critical step towards the development of VR systems to enhance human capabilities in various contexts.

An association between factors that appears quite straightforward is one between depth representation or pictorial realism and presence in virtual environments. Depth perception has been acknowledged as the “missing link” between pictorial realism of VR simulations and presence. The expectation about such association dates back to the early Renaissance period, when perspective renderings painted on walls at natural scale caused a strong perceptual response where the viewer would feel transported into the virtual space created by the painting (Bertol, 1997; Brooks, 2017). Thus, it appears that the link between depth representation (perspective) and presence is, again, 3D perception: depth representation would enable better 3D perception, which in turn would enable greater levels of presence (Figure 7). Thus, one must perceive the three-dimensional space portrayed in a representation in order to perceive oneself in the place depicted. It should be noted that in Figure 7, presence occupies the same position as realism of representation in Figure 3, in a deliberate assumption (based on the literature) that in virtual environments presence and realism are products of similar processes. That is, 3D perception is assumed to determine simultaneously the realism of simulations and presence in virtual environments.

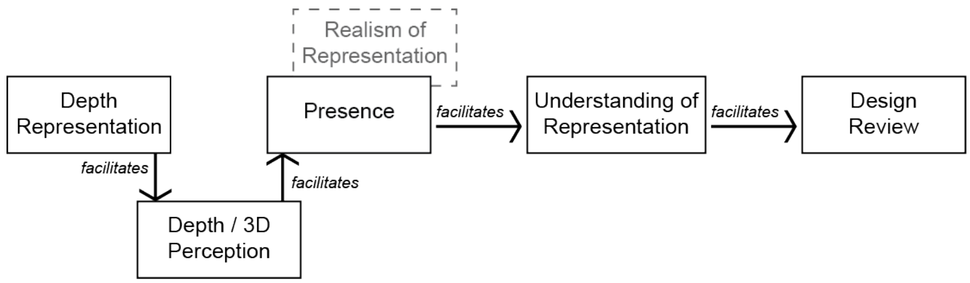


Figure 7. Relationship between depth representation and presence

Contemporary theorists reinforce the expectation of a direct association between presence and 3D perception. Bertol (1997) suggests that immersive environments improve spatial perception (presence – perception); Steuer (1992) states that immersive environments enhance the level of presence due to, among other factors, stereoscopic visualization (depth representation – perception – presence); Witmer and Singer (1998) state that presence derives from perceptual processes, which in turn are subject to realism of a simulation (depth representation – perception – presence). Combined, these arguments corroborate to the expectation of a positive association between 3D perception and presence in virtual environments.

If true, this knowledge could contribute to the development of VR systems in which depth representation is manipulated to enhance presence levels. VR developers may focus on providing accurate 3D perception (through accurate depth representation) in order to deliver highly engaging and immersive experiences, which can benefit learning (Oren et al., 2012) and task performance (Faas et al., 2014).

Zikic (2007) stated that there is not much information on the requisites for spatial perception of virtual environments, as well as no clear evidence about its relationship with sense of presence. A few empirical studies investigated the relationship between presence and distance perception in virtual environments (e.g., Thompson et al., 2004; Kalisperis et al., 2006; Interrante et al., 2006; Renner et al., 2013) and showed that accurate distance estimation may not be necessarily an evidence of great levels of presence. People may still perform well in estimating distances in virtual environments that do not offer the conditions for great levels of presence (Interrante et al., 2008).

In the quest for the identification of factors that promote presence and of the relationship between presence and perception in virtual environments, this question aims to verify whether presence and 3D perception are associated within each VR system (non-immersive and immersive). The expected association tested is unidirectional, that is, 3D perception affecting presence. It is important to emphasize that due to the nature and design of this study cause-effect relationships are not detectable, only associations.

CHAPTER 4

METHODOLOGY

4.1 Past Experiments

Spatial perception (and related terms) and presence in virtual environments remains a relatively unexplored niche of research. Methods of data collection are difficult to develop since variables of interest often result from intricate and eventually unclear cognitive phenomena. Nonetheless, identifying and characterizing technological and human factors that affect presence and perception in virtual environments has been acknowledged by many scholars as one of the most critical steps for the development of VR systems to enhance human capabilities in various contexts (Slater, 1999; Zikic, 2007; Interrante et al., 2008). Faas et al. (2014) provide a thorough review of past attempts, most of which looked into the relationship between presence or spatial perception and possible underlying factors (VR system's properties and user-related factors).

4.1.1 Measuring Presence

Measurements of the psychological state of presence in virtual environments are quite difficult to perform since presence is a subjective sensation not easily available to objective observation. Witmer and Singer (1998) were pioneers and developed the first Presence Questionnaire (PQ). Assuming that presence was a function of a person's response to the VR system's properties, their questionnaire aimed at providing presence measurements based on users' opinions on the extent to which certain system characteristics led them to experience presence. Most questions were derived from involvement and immersion factors assumed to underlie presence, including sensory, control, distraction, and realism factors. Their selfassessment of the PQ scale granted "credence to the factors that were used in generating the PQ scale and to the scale structure", demonstrating internal consistency, reliability, and validity of the instrument, although some of those factors were not thoroughly checked. In addition, Witmer and Singer (1998) believed that measurement methods should also account for "individual tendencies" that reflect one's inclination to become involved and immersed in the virtual experience. Witmer and Singer (1998) found that these individual tendencies – measured by their Immersive Tendencies Questionnaire (ITQ) – are able to predict level of presence, although the correlation between ITQ scores and levels of presence reported was weak.

The major critics to Witmer and Singer's Presence Questionnaire (e.g., Slater, 1999; Usoh et al., 2000; Faas et al., 2014) argue that while the authors acknowledge that presence derives from perceptual processes (involvement and immersion), their instrument may not measure the psychological state of presence but a person's subjective opinions about various properties of a VR system. Indeed, questions like "*How much did the visual aspects of the environment involve you?*" or "*How natural was the mechanism which controlled movement through the environment?*" (Witmer and Singer, 1998) appear to relate more to the characteristics of the system than to a user's cognitive experience. Nonetheless, presence and spatial perception questionnaires have been adapted from Witmer and Singer's instruments and used in various studies in the built environment domain (e.g., Ruschel et al., 2005; Kalisperis et al., 2006; Zikic, 2007; Castronovo et al., 2013; Heydarian et al., 2015b; Paes and Irizarry, 2018; Khashe et al., 2018).

A few among those studies sought to compare presence levels across different VR systems. Khashe et al. (2018) conducted a between-subject study to examine the effects of immersive and non-immersive VR platforms on compliance with environmental requests, reading performance, presence and motion sickness. They also analyzed the influence of some individual factors on the dependent variables, as well as the interaction between presence and other dependent variables within each condition. In a within-subject study, Castronovo et al. (2017)

investigated the relationship between presence and VR systems in design review. Presence was given by participants' self-reported experiences of realism of movement, "physical" presence, attention, and their sense of "being part of the virtual environment", among others. They utilized different simulations in each VR condition (Revit-based *vs.* Unity3D-based simulations). Ozcelik and Becerik (2018) designed a within-subject study to investigate the relationship between presence and potential influencing factors (comfort, satisfaction, number/type of interactions, perceived temperature) and found a positive correlation between perceived thermal comfort and presence when analyzing data from all conditions combined. Higuera-Trujillo et al. (2017) conducted a thorough between-subject experiment to compare the differences in psychological, physiological, and presence responses among different display formats including photographs, 360° panoramas, and a virtual environment. Responses from display modes were "standardized" over physical-world responses beforehand to simplify the comparisons among display types. They also adopted the SUS presence questionnaire (Usoh et al., 2000). Naturally, presence responses were not collected in the physical environment condition. Results indicated that VR offers the closest-to-reality experience with respect to the user's physiological responses, and that physiological and psychological responses correlate with the sense of presence.

4.1.2 Measuring Three-dimensional Perception

While there are many methods to assess a person's spatial perception in virtual environments (Interrante et al., 2006), performing such experiments correctly is difficult and care should be taken to understand the factors involved (Gooch and Willemsen, 2002). Cognitive psychologists have conducted the most impactful investigations thus far, although studies from other disciplines also provide relevant insights. In general, due to their direct association, distance estimation has been widely used as a metrics of depth/three-dimensional perception in both physical and virtual environments (e.g., Witmer and Sadowski, 1998; Sinai et al., 1999; Gooch and Willemsen, 2002; Thompson et al., 2004; Interrante et al., 2006; Renner et al., 2013). In studies that compare distance estimation between virtual and real-world settings, accuracy of virtual distance estimation is usually given by the percentage of the absolute/actual distance estimated by a participant (Ziemer et al., 2009). However, when a study compares distance estimation between different virtual systems, accuracy of estimation is provided by the deviation of a participant's estimates in the virtual environment with respect to her/his estimates in the real world (Higuera-Trujillo et al., 2017; Paes et al., 2017).

Distance estimation can be performed through different methods. Most studies adopt egocentric judgment techniques, i.e., distance estimation from the observer at a fixed position to a target object. Alternatively, distance estimation can be based on the time it takes a participant to walk towards target-objects, also referred to as "time-to-walk" estimates (Ziemer et al., 2009).

Findings suggest that people are very good in estimating distances in the real world. While egocentric distance estimation within action space (up to 30 meters radius) is quite accurate (Ziemer et al., 2009), people tend to underestimate values over greater distances. In general, egocentric judgments are approximately 8% underestimated (Witmer and Sadowski, 1998), that is, they do not correspond to the absolute distances, the actual measures (Gooch and Willemsen, 2002). Interobject distance judgments are also underestimated. However, when walkthrough exploration is allowed, distance perception is near veridical (Cutting and Vishton, 1995; Sinai et al., 1999; Gooch and Willemsen, 2002). When exploration is allowed in interobject distance estimation, a participant might simply position oneself near one of the objects and look at the other one, which approximates interobject estimation to egocentric estimation. In virtual environments, egocentric distance estimation is even less accurate than those 8% of underestimation in the real world: virtual dimensions appear approximately 15% shorter than the actual/absolute dimensions (Thompson et al., 2004). While walkthrough exploration may benefit egocentric distance estimation to virtual targets (Thompson et al., 2004), in general, distances appear more compressed in virtual environments than they do in the real world (Ziemer et al., 2009).

Virtual environments may evoke similar responses to those observed in physical environments (Higuera-Trujillo et al., 2017). However, apparently, this is not entirely true for distance perception responses. As discussed above, past studies that have compared distance judgments in the real world to judgments in a virtual environment have almost consistently shown that egocentric distances are underestimated in virtual environments in relation to egocentric estimates in the real world (Interrante et al., 2006; Renner et al., 2013). Particularly large differences were observed between distance judgments in the real and virtual conditions by Witmer and Sadowski (1998).

In summary, people tend to underestimate egocentric distances in both physical and virtual environments, but the underestimation seems larger in the virtual setting. The factors that account for the systematic and large egocentric underestimations in virtual settings remain unknown. If egocentric distance judgments in the virtual world are not as good as in the real world, according to Cutting and Vishton (1995) one could assume that there are not enough depth cues available in virtual environments – e.g., stereopsis – to enable a realistic depth perception (close to perception in the physical world).

As per Kimura et al. (2017), visual spatial cues can be broadly categorized into geometric (e.g., distance or direction) and featural cues (e.g., color or texture). The latter may not significantly impact distance perception. Evidence shows that the quality of graphics (rendering effects associated with featural cues) of the virtual simulation has little to no effect on distance judgments. A thorough study conducted by Thompson et al. (2004) found that egocentric distance judgments based on wireframe renderings (with floor tiling grid) are just as good as judgments from 360° photographs of the actual environment presented with the same display system. The floor tiling grid pattern in the low-fidelity model seems to have provided critical cues for depth perception, facilitating distance estimation (Sinai et al., 1999). Alternatively, other studies suggest that stereopsis, geometrical floor pattern, and high-fidelity graphics are all possibly critical cues for distance perception (Renner et al., 2013).

4.2 Research Methodology

The research method is based on previous research on visual perception and sense of presence in virtual environments that adopted presence and/or visual perception questionnaires to assess user experience (e.g., Witmer and Singer, 1998; Usoh et al., 2000; Ruschel et al., 2005; Kalisperis et al., 2006; Zikic, 2007; Castronovo et al., 2013; Faas et al., 2014; Heydarian et al., 2015b; Paes et al., 2017; Paes and Irizarry, 2018). It consists of four methodological aspects, as described below:

a) *Context*: As per Sacks et al. (2013), IVR effectiveness has not been rigorously tested in building construction yet. In this research, effectiveness relates to a specific context of technology usage within the construction industry, namely, the design review process (Fernando et al., 2013).

b) *Comparative*: In this study, effectiveness of VR systems refers to the degree to which such systems are successful in providing users with accurate 3D perception and high levels of presence. A VR system could only be said *more* effective in comparison to something else. That is why, as opposed to self-evaluations and inspection-based assessments, comparative studies facilitate the observation of improvements – as these can only be established against something else – providing a means for determining the level of effectiveness (Kim et al., 2013). Thus, this study compares an immersive VR system to a non-immersive VR system.

c) *Experimental and quantitative*: Experimental research appears to be the only approach that enables researchers to make judgments about beliefs and assumptions with systematically measured confidence and reliability (Lazar et al., 2017). This study utilizes quantitative data analysis methods including inferential statistics, allowing for inference and generalization of results across the user population and context of application.

d) *Approach*: User-centered investigations are critical to ensure that the technology develops with adequate concern for its users (Stanney et al., 1998). In this study, the effectiveness of the IVR system is evaluated on the basis of users' cognitive responses, i.e., their perception of the three-dimensional relationships of a virtual environment and levels of presence reported.

The research methodology comprises the steps listed below, which are described in details throughout this Chapter.

1. Definition of statistical hypotheses (Chapter 3);
2. Development of experimental design;
3. Setup of experimental conditions;
4. Development of research instruments;
5. Development of research protocol;
6. Pilot experiment;
7. Sampling;
8. Recruitment of participants;
9. Assignment of participants to experimental conditions;
10. Data collection;
11. Data analysis (Chapter 5).

4.2.1 Experimental Design

This is an empirical, quantitative, relational study, although it carries many characteristics of experimental research (such as an experiment), as most Human-Computer Interaction (HCI) studies. Relational investigations allow for the identification of connections between multiple factors. If results of inferential statistical analyses are significant, it suggests that the relationships are true. However, relational research can rarely determine the causal effect between the variables (the causes of the observed relationships), as cause-effect relationships could be due to unappreciated variables (Lazar et al., 2017).

The method is formally defined as a controlled experiment utilizing survey questionnaires to collect user experiences. Experiments are studies that involve multiple groups or conditions to which participants are randomly assigned. An experiment has the following characteristics: a) it involves at least one testable hypothesis and aims to validate it, b) it involves at least two conditions or groups (a single independent variable could assume different values in each condition, such as “medicament *vs.* no medicament”), c) dependent variables are normally measured through quantitative measurements, d) results are analyzed through statistical significance tests, e) it is designed and conducted with the goal of removing potential biases, and f) it should be replicable in different circumstances (Lazar et al., 2017).

This study adopts a within-group experimental design, also referred to as “within-subject” design, where participants are exposed to all experimental conditions. Data analysis consists of comparing the performances of the same participants under different conditions. Because 3D perception and presence involve significant cognitive functions, individual differences are expected to largely affect the outcomes. Such individual differences are better controlled in a within-subject design, which excludes the variance between subjects due to those differences in the comparison of effects of different conditions, since each participant serves as her/his own equivalent in the comparison (Lazar et al., 2017). Consequently, power to detect existing differences is usually much higher in within-subject designs than in between-subject designs given the same sample size (Thompson and Campbell, 2004; Charness et al., 2012). Nonetheless, within-subject experiments must be carefully designed to ensure that the benefits of such a design (smaller sample size, greater power, etc.) outweigh potential drawbacks (practice/learning/carry-over, fatigue, and expectancy effects) (Thompson and Campbell, 2004).

In this study, the number and values of the independent variable (a single independent variable of two different values) create two different experimental treatments or conditions, namely, the "niVR" (a BIM model displayed through a laptop screen) and the "IVR" (an immersive virtual reality system) visualization modes. A participant's performance in 3D perception and level of presence (dependent variables) are collected in each condition. It should be noted that a user's performance in 3D perception is determined – or "standardized" (Higuera-Trujillo et al., 2017) – against her/his performance in the physical environment (PhE). The experimental design structure is provided below (Figure 8).

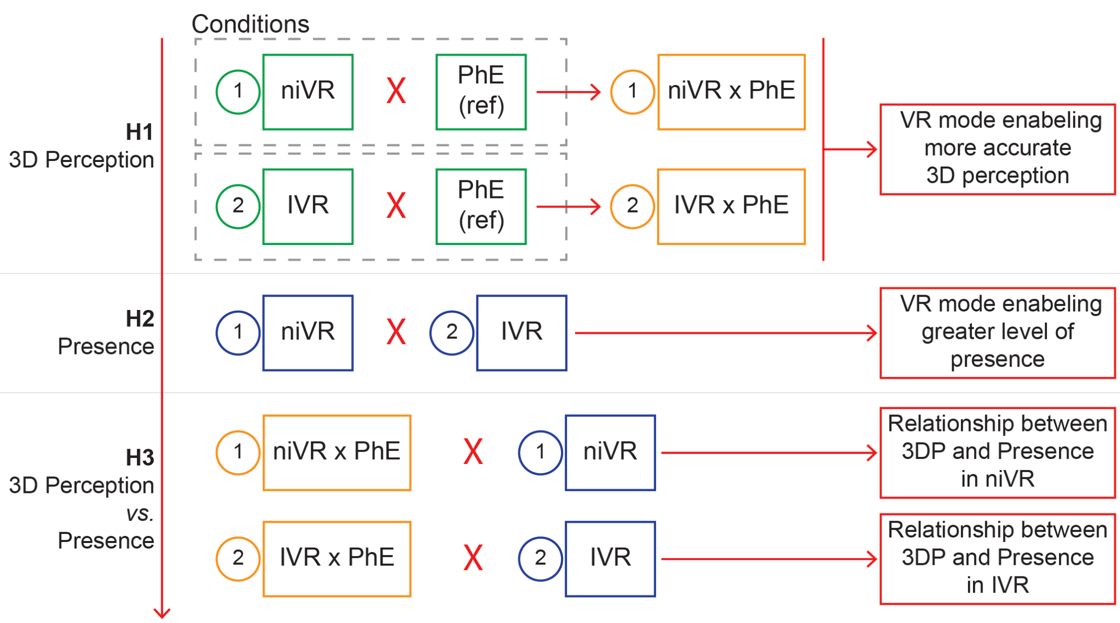


Figure 8. Experimental design structure

The dependent variables of 3D perception and presence could be affected by several factors that studies of this nature cannot fully control. These are known as random or confounding variables (Pourhoseingholi et al., 2012), or yet, covariables (Lazar et al., 2017). Participants could perform better in 3D perception and report greater levels of presence due to those factors. In this study, those consist of individual factors of age, gender, educational level, bachelor's major, current major, experience in design review, computer usage, experience with 3D virtual environments, familiarity with the experiment environment, and spatial ability.

There are various ways to exclude or control confounding variables. Their effects can be controlled by choosing appropriate experimental design (Thompson and Campbell, 2004) and/or in data analysis through statistical methods (Pourhoseingholi et al., 2012). In a within-subject experimental design, these variables are almost fully controlled because a participant's performance in a condition is compared against her/his own performance in the other condition such that individual factors are equally impactful in both conditions. Although familiarity with the experiment environment is not expected to affect 3D perception and presence (Paes et al., 2017) it is included among this study's confounders. Regardless of the unlikelihood of impacts of confounders on the dependent variables due to the experimental design, their effects will be verified in data analysis to check for the effectiveness of measures taken to control them with the experimental design.

The experiment allows for a quantitative comparison of responses between two distinct visualization modes: 1) a BIM model displayed through a laptop screen (non-immersive VR), and 2) a BIM model viewed through a commercial head-mounted display (immersive VR). Data collection makes use of various survey questionnaires. From the statistical comparison of responses, it is possible to identify and quantify differences in 3D perception and presence between conditions.

4.2.2 Experimental Conditions

The lobby of the Caddell Building on the Georgia Tech campus was selected as the experiment environment (Figure 9). The criteria for its selection involved ease of access to information required to build a BIM model of the physical location (CAD drawings, on-site checking of as-built dimensions), ease of access to the physical location, physical proximity to the room/laboratory where the first part of the experiment session would take place, moderate level of architectural complexity, diversity and characteristics of constructive elements.

Experiment sessions took place in the lobby and in room 208 of the same building. The reference mode is the actual lobby, whereas visualization modes are virtual environments, as described next.

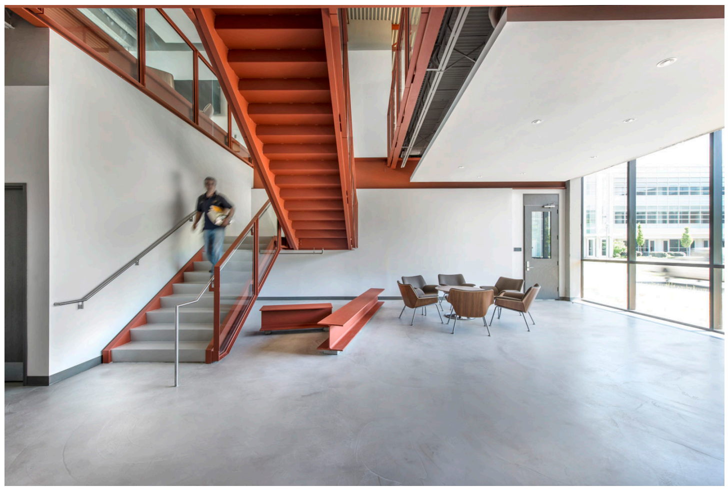


Figure 9. Caddell Building lobby

The decision between using a high- or low-fidelity model in the VR modes is guided by the general research motivation: to identify benefits of immersive visualization of BIM models. As pointed out by Berg and Vance (2016b) and Paes and Irizarry (2019), the relevance of pictorial realism in virtual environments is strictly a function of the questions being explored. High-fidelity simulations are valuable but not always required for decision-making in the architectural design process. For designers who concentrate on the fit, form, and function of a space, a model's geometry must be accurate and representative of the design solutions with respect to scale, size, orientation, and position. In this scenario, decision makers are less interested in the pictorial realism of the simulation and more interested in whether the design fulfills the technical specifications. Thus, using high-fidelity renderings may not be a priority in this case. Indeed, low-fidelity BIM models are the most widely used simulations in design review and across the construction industry (Eastman et al., 2008).

Therefore, the use of a BIM model of low-fidelity graphics in both VR modes was set as a first requirement. Regardless of the controversy about the extent to which the quality of graphics affects perception and presence in virtual environments, featural spatial cues (Kimura et al., 2017) should be kept the same across VR modes to allow for adequate observation of display effects (stereopsis, field of view, and interactivity effects). Keeping graphics quality the same across VR conditions eliminates the effects due to differences of graphics and automatically leaves the effects due to the visualization technology "at their will."

Thus, both VR conditions feature: a) first person view (FPV) walkthrough mode of navigation, b) level of graphics quality equivalent to the quality currently provided by BIM software applications, c) identical levels of graphics quality, and d) identical user interfaces (UIs). User interface is deemed an impactful component of visual stimuli in VR conditions, besides the projections of the virtual model itself. Different UIs across VR conditions would provide different resulting visual stimuli in each condition. Peripheral distracting noise such as menu bars on UIs of most BIM applications is not desirable as well. In regards to the level of graphics, when converter applications are used to bring BIM models into HMDs, these models usually undergo some light rendering work automatically (reflections, shadows, and textures are modified or added), resulting in a model with slightly different pictorial parameters from the original BIM model. Nonetheless, the resulting model is quite like the original, with low-fidelity graphics. Using the same converter application in the non-immersive VR condition ensures similar and equally free-of-noise UIs, as well as identical levels of graphics quality. The decision on what BIM application to adopt was subject to what converter application would be used next to provide virtual experiences in both VR conditions. The characteristics of the applications tested are summarized in Table 3 next.

Table 3. Converter applications

BIM applications supported	Converter application	VRpc + IVR support?	First person walkthrough feature?	Available License
SketchUp	Yulio	Yes	No	Free 30-days trial
Revit, ArchiCAD	Enscape	Yes (high-fidelity only)	Yes	Free educational, semester-long
Revit, ArchiCAD	Revizto	Yes	Yes	Paid
SketchUp, Revit	Kubity	Yes	Yes ("hidden" edges)	Free
SketchUp, Revit	Prospect	Yes	Yes	Free educational, year-long

As shown in Table 3, only Enscape, Revizto, Kubity, and Prospect support BIM applications (Revit and ArchiCAD) and feature FPV walkthrough mode of navigation simultaneously. Only Enscape, Kubity, and Prospect offer educational licenses for research use (upon request). In Kubity, the elements' edges fade out during head movement, which could affect visual perception. Enscape can only convert a BIM model into a high-fidelity simulation (high-end graphics) much superior than in any BIM software application, and does not provide the option to convert into low-fidelity simulations. Consequently, for the purposes of this study, IrisVR Prospect Plus was chosen as the application to run an Autodesk Revit model in both niVR and IVR conditions.

Both simulations are formally classified as *exploratory simplified virtual reality*. The term *exploratory* refers to when a user can perform visual search, exploring the virtual environment at her/his will, defining her/his own path, stopping at desired locations and focusing on certain objects. The term *simplified* refers to the degree of pictorial realism and vividness of a virtual environment – its graphics quality (Bertol, 1997; Ruschel et al., 2005). In this study, both VR modes are low-fidelity simulations with limited rendering effects (lighting, coloring, and textures). In summary, they only differ in interaction devices (keyboard and mouse *vs.* wireless controllers) and display type (laptop monitor *vs.* head-mounted display). Figure 10 shows a set of screenshots of the virtual environment in both VR modes.

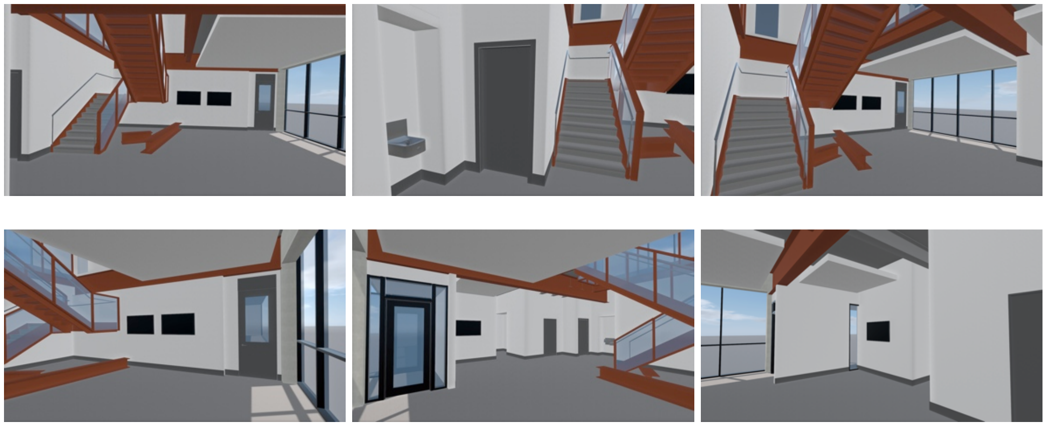


Figure 10. Screenshots of the virtual environment

4.2.2.1 Setup of VR Modes – Apparatus

For the niVR mode, the Revit model was exported to Prospect converter application. Hardware consisted of a conventional workstation comprising a high-performance 15” laptop, keyboard and mouse (as interaction devices). Detailed information on the hardware used is provided in Table 4 below. The laptop was placed on a conventional office table so that users could navigate through the virtual environment while seated, using the interaction devices.

Table 4. niVR Hardware

Item	Sub item	Specification
Computer	Manufacturer	Dell
	Model	Alienware 15
	Operational System	Windows 10 Home
	CPU	7th Generation Intel® Core™ i7-7700HQ (Quad-Core, 6MB Cache, up to 3.8GHz w/ Turbo Boost)
	GPU	NVIDIA® GeForce® GTX 1070 with 8GB GDDR5 Overclocked
	Memory	32GB DDR4 at 2400MHz (2x16GB)
	Storage	256GB PCIe SSD (Boot) + 1TB 7200RPM SATA 6Gb/s (Storage)
	Display	Built-in 15.6” FHD (1920 x 1080) 60Hz IPS Anti-Glare 300-nits NVIDIA G-SYNC Enabled, Non-touch
	Interaction devices	Built-in keyboard + standard mouse

For the IVR mode, the Revit model was exported to Prospect converter application, which automatically exports the model into the head-mounted display (HTC Vive™). Hardware consisted of the same conventional workstation comprising a high-performance 15” laptop and 2 wireless controllers (as interaction devices). The HMD was set up in room-scale mode so that users could navigate through the model in standing position, with limited space for moving around the experiment room (approximately one step in each direction). Base stations were installed facing each other, at a distance of approximately 13 feet (4 meters), and at 6 feet height (1.8 meter). Detailed information on the hardware used is provided in Table 5 below.

Table 5. IVR Hardware

Item	Sub item	Specification
Computer	Table 4	
HMD system	Manufacturer	HTC
	Model	VIVE VR System
	Display	Screen: Dual AMOLED 3.6” diagonal Resolution: 1080 x 1200 pixels per eye (2160 x 1200 pixels combined) Refresh rate: 90 Hz Field of view: 110 degrees Eye Relief: Interpupillary distance and lens distance adjustment
	Interaction devices	2 wireless controllers with haptic feedback
	Sensors	2 base stations (movement tracking sensors), G-sensor, gyroscope, proximity

4.2.3 Instruments

This section provides a detailed description of metrics and related instruments developed and used in this research.

1. Consent Form – Written consent is obtained at the beginning of the experiment session by the researcher in charge of the session. The Consent Form is handed to the participant after being introduced to the purposes and procedures of the study, and after the researcher makes sure the participant has had all questions answered. A copy of the Consent Form is provided in Appendix A.

2. Demographic Questionnaire (DQ) – Used to collect demographic data of participants, including information on individual characteristics that could possibly affect 3D perception and presence. It is administered immediately after obtainment of consent signature. A copy of the DQ is provided in Appendix B.

3. Spatial Ability Test (Revised PSVT:R) – The spatial ability form, namely, the Revised Purdue Spatial Visualization Test – Visualization by Rotations (Revised PSVT:R), is a well-accepted multiple-choice test used to collect human spatial ability (Yoon, 2011). Previous studies suggest that spatial ability is positively correlated with users' performance in spatial tasks (Kovac, 1989), retention and achievement in science disciplines, and may improve with training (Kinsey et al., 2006). In a study conducted by Oren et al. (2012), participants completed a spatial ability test so the researchers could control for variance of learning performance due to spatial ability differences. An adapted version of the Revised PSVT:R is administered following the completion of the Demographic Questionnaire. A copy of the adapted Revised PSVT:R is provided in Appendix C. Spatial ability is expected to have little impact on the dependent variables. The assumption is that even those who demonstrate low spatial ability should be able to achieve great levels of presence (Witmer and Singer, 1998) and to perform well in spatial perception tasks (Calderon-Hernandez et al., 2019).

4. 3D Perception Questionnaire (3DPQ) – Used to collect participants' three-dimensional perception (horizontal, vertical, and depth judgments combined) of an architectural design in each VR mode (niVR and IVR) as well as in the reference mode (PhE). The 3DPQ comprises twelve objective questions that prompt the respondents to estimate egocentric distances to objects in space (egocentric distance estimation) and distances between objects in space (interobject distance estimation) (Table 6). The questionnaire was based on previous studies that have developed and used similar instruments (e.g., Kalisperis et al., 2006; Zikic, 2007; Paes et al., 2017).

In order to ensure its qualitative validity, the development of questions involved consultation with Mr. William T. Oswell, chief architect in the Facilities Design and Construction Department of Georgia Tech (Figure 11), three other professional architects, and three professional civil engineers, who provided general insights in the determination of the three-dimensional information of interest in collaborative design review, that is, the spatial relationships in a project's 3D model that would be of interest to professionals involved, providing support to reasoning and feedback. For example, the structure team may be interested in the layout and relative size of structural elements and spans proposed by the architecture team (identifying and interpreting such visual information requires three-dimensional perception). These requisites were translated into 3DPQ's questions so that they could better address what the design review team seeks over review meetings using 3D models.

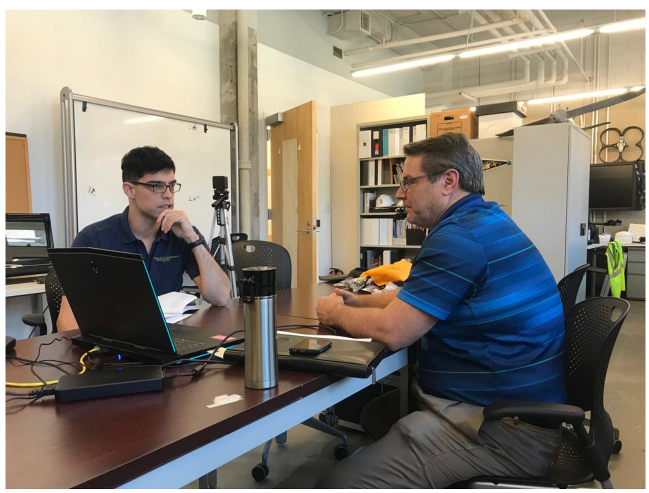


Figure 11. Consultation meeting with architect during the development of the 3DPQ

Questions are multiple-choice, with objective alternative options comprising intervals of estimates (e.g., up to 10m, up to 12m, and so on.) in lieu of Likert-scale "level-of-agreement" alternatives, in an attempt to mitigate subjectivity in measurements of inherently subjective user experiences such as visual perception.

Table 6. 3DPQ's questions categories

Category	Dimension	Exploration	Question #
Egocentric distance estimation	Depth	Not allowed (fixed position)	1, 2, 3
Interobject distance estimation	Depth	Not allowed (fixed position)	4, 5, 6
	Horizontal	Allowed	7, 8, 9
	Vertical	Allowed	10, 11, 12

Randomization software was used to sort the 3DPQ's questions in order to minimize learning effects, originating three unique 3DPQs (all 3DPQs comprise the same questions, but in different sequences). The 3DPQs are administered during walkthroughs in the virtual environments (questions and alternative options are read out loud) and then in the visit to the physical environment. A copy of the 3DPQ-key is provided in Appendix D.

5. Presence Questionnaire (PQ) – Used to collect the level of presence a participant experienced during the 3D perception tasks in the virtual environments (niVR and IVR). It is administered immediately after completion of the virtual walkthrough. A copy of the PQ is provided in Appendix E.

The PQ developed and used in this research is based on many attempts of measuring presence in virtual environments, both in cognitive psychology and construction fields (Witmer and Singer, 1998; Usoh et al., 2000; Kalisperis et al., 2006; Zikic, 2007; Faas et al., 2014; Heydarian et al., 2015b). It comprises an adapted collection of questions from instruments of Witmer and Singer (1998), Usoh et al. (2000), and Zikic (2007), as well as one new question (question 7), and a final question on motion sickness, in a total of eleven items. It was mainly based on the instrument developed by Usoh et al. (2000) named SUS (Slater-Usoh-Steed). Furthermore, it adopts a 7-point Likert scale to maintain consistency with the instruments of Witmer and Singer (1998) and Usoh et al. (2000). The decision of using the SUS as main reference was informed by previous studies (Faas et al., 2014) showing that measures of presence have been more successful when metrics and respective instruments address the three aspects of presence defined by Slater (1999): 1) the sense of being in the virtual environment, 2) the degree to which the virtual environment becomes the dominant reality, and 3) the extent to which the virtual environment is remembered as a 'place'. Usoh's et al. (2000) and Slater's (1999) approaches appear to better address presence as a psychological state resulting from virtual input rather than as a direct function of system properties, such as argued by many scholars regarding Witmer and Singer's (1998) presence questionnaire.

Table 7 presents a summary of variables and instruments used in the VR conditions and in the physical environment.

Table 7. Variables and instruments per visualization and reference modes

	niVR	IVR	PhE (reference mode)
Variables	3D Perception	3D Perception	3D Perception
	Presence	Presence	
Instruments	3DPQ-niVR	3DPQ-IVR	3DPQ-PhE
	PQ	PQ	

4.2.4 Research Protocol

At Georgia Tech, Institutional Review Board (IRB) approval is required in advance for all research projects that involve human subjects. The Central IRB reviews most human subject research activities and has the authority to approve, require modifications, or disapprove projects that fall within their jurisdictions as specified by both federal regulations and Georgia Tech policies and procedures. A research protocol was developed and submitted to IRB review on September 4, 2018 (protocol H18334). It was approved by the Central IRB on September 11, 2018, with no modifications required. On its expiration date, September 10, 2019, it was renewed for one additional year. A copy of the protocol is provided in Appendix F.

4.2.5 Pilot Experiment and Experiment Simulation

Following the development of the experimental design and IRB approval, a pilot experiment session was conducted to check for the adequacy of the experimental design, data collection instruments, procedures, and equipment. This phase is particularly important to review and perform final adjustments to the experimental procedures and survey questions. The pilot experiment was performed on September 28, 2018. Issues involving the IVR platform setup and software settings for optimal user experience, duration of experiment session, phrasing and structure of survey questions were addressed.

Following the pilot experiment session and necessary modifications, an experiment simulation was conducted over the Fall semester of 2018 (October 1-17) with the participation of 17 people. The experiment simulation sessions provided data on the effectiveness and adequacy of the research method, experimental procedures and instruments. The results from these trials were presented to and thoroughly reviewed by the Dissertation Committee Members at and following the Proposal Defense on November 29, 2018. Several modifications to the research methodology were performed, as follows: a) modifications to the BIM model utilized in the virtual environments, b) decrease in the number of experimental conditions and survey forms, c) decrease in the number of dependent variables, d) modifications to some of the survey questions, e) addition of new questions, and f) other cosmetic modifications to the experimental procedure.

4.2.6 Sampling

On probabilistic or random sampling, the goal is to achieve a population estimate, i.e., a sample made of random selection of members from the entire population (which is rarely possible). In HCI research, however, population estimates are generally not the goal. In this study, participants are recruited in a nonprobabilistic manner (also referred to as nonrandom convenient sampling). The nonprobabilistic sample is made up of volunteers, that is, people who accepted to participate voluntarily by invitation. Basic demographic data of research participants can give the representativeness of nonprobability-based samples and validity of survey responses (Lazar et al., 2017).

Although many scholars in the social sciences and statistics argue that without strict probabilistic sampling no survey data are valid, HCI academics have a long history of using convenient sampling. Small, nonprobabilistic samples are used throughout HCI research on a regular basis, without concern. Part of this difference may be due to the different nature of research between those fields. In HCI, researchers typically collect the data themselves as opposed to researchers in statistics or social sciences, who often refer to already existing probability-sampled data across open-source repositories. Nonprobability-sampled studies may be the most natural data collection method for investigating new usage phenomena; if no data exists about user experience with a certain technology, nonprobabilistic sampling is applicable (Lazar et al., 2017).

The population of interest consists of the body of industry workforce (current and future workforce) traditionally involved in collaborative design review, including, but not limited to: architecture, civil engineering, and building construction students, professional architects, civil engineers, construction and facility managers, and trade contractors. This study investigates the effects of different treatments on that specific population – it is not concerned with the effects on the general population. Thus, results from the experiments would have implications on that particular population only.

As per Lazar et al. (2017), in HCI research 30 respondents would be considered a baseline minimum number of participants. Previous studies provide an overall idea of sample sizes. The sample size in the study conducted by Faas et al. (2014) was of 30 participants. Sacks et al. (2013) conducted between-group experiments in which the total sample sizes were of 20 and 25 participants (divided into two groups of 10-12 participants each). Interrante et al. (2006)

conducted two experiments for which they recruited 7 and 10 participants. Witmer and Singer (1998) performed a set of experiments to validate their instruments, having recruited 38 participants on average in each study.

4.2.6.1 Power Analysis

An *a priori* power analysis (prior to collecting data) should be conducted when a researcher is planning a study and wants to determine the power of a statistical test given a sample size if differences in the response variable between conditions were similar to estimated values (effect size). Power is the probability that a particular statistical test will detect a difference when it exists, thus allowing for the correct rejection of the null hypothesis (Green and MacLeod, 2016). Therefore, *a priori* power analysis provides the adequate sample size to answer the research questions.

On one hand, by enrolling too few participants a study may not have enough statistical power to detect existing differences (type II error: not detecting a difference when it exists). On the other hand, a study might be overpowered with an excessively large sample, consuming more resources than necessary. Adopting a within-subject design contributes to increasing statistical power since each participant serves as her/his own equivalent hence excluding the variance between subjects due to individual differences in the comparison of effects of different conditions. Studies show that power is much higher in within-subject designs than in between-subject designs given the same sample size (Thompson and Campbell, 2004; Charness et al., 2012).

In general, power is determined by the following:

- Sample size / number of observations (positive relationship). It should be noted that number of observations could be given by number of participants, number of conditions, and number of questions (responses from each participant).

- Alpha (α) – Also referred to as significance level, it is the probability of a type I error, i.e., of detecting a difference when it does not exist, or yet, of mistakenly rejecting the null hypothesis when it is true. Also referred to as confidence level, $1 - \alpha$ is the probability of correctly retaining the null hypothesis (when it is true). Traditionally, studies in the field adopt a α value of 0.05, defining a 5% chance that a significant difference is actually due to chance and not a real difference.

- Beta (β) – The probability of a type II error, i.e., of not detecting a difference when it exists, or yet, of mistakenly retaining the null hypothesis when it is false. Also referred to as a study's statistical power, $1 - \beta$ is the probability of correctly rejecting the null hypothesis (when it is false) – which is what studies usually aim at. Studies in the field adopt a β value of 0.2, indicating a 20% chance that a significant difference is missed. Researchers usually use a 4-to-1 trade off between β and α to get an 80% power (Gheisari, 2013). That is, assuming a α level of 0.05, β would equal to 0.2, resulting in a power ($1 - \beta$) of 0.8.

- Effect Size – The estimated magnitude of the effect of a treatment, i.e., the size of the difference between conditions. A large sample size increases the power of detecting small-sized effects, i.e., significant small differences between conditions, due to shrinkage of confidence intervals (Thompson and Campbell, 2004). Given a fixed sample size, power generally increases with effect size, with larger effects being easier to detect (Green and MacLeod, 2016).

As pointed out by Gheisari (2013), HCI studies usually verify whether the observed difference is real or random, the magnitude of the difference, and the meaningfulness of such magnitude. Therefore, not only performing significance tests but also defining a meaningful effect size is important to ensure the relevance of results. An ideal practically significant result would not only be statistically significant but also have a meaningful effect. The term *practical significance* implies a research result that will be viewed as having importance for the practice. There is no specific statistical test that directly measures the practical significance of effects observed (Gheisari, 2013); it can only be given by defining an adequate effect size in light of the research goals. Naturally, meaningfulness of an effect magnitude is subjective and varies across studies. While Cohen (1988) provides standard values of effect magnitudes for studies in the social and behavioral sciences, effect sizes should be determined on a case-by-case basis, as it is subject to the response variables, aims, instruments, and other specific aspects of a study. Frequently, large-sized effects are chosen to ensure practical significance of results (Gheisari, 2013).

In this study, *a priori* power analysis utilizes the SIMR R package, which can calculate power for Generalized Linear Mixed Models (GLMM) using Monte Carlo simulations (Green and MacLeod, 2016). GLMM are suitable for the statistical analysis of count (number of occurrences), categorical (number of observations falling into separate categories), proportions, and continuous data, and are used in this study for hypothesis testing given the diverse characteristics of the data collected. Thus, power analysis should be conducted based on GLMM, which are the methods adopted for statistical data analysis. The SIMR runs a power analysis given a statistical model and experimental design, and calculates power curves to assess trade-offs between power and sample size. In SIMR, power is calculated by repeating the following three steps: a) simulate a new value for the response variable using the model provided, b) refit the model to the simulated response, and c) apply a statistical test to the simulated fit. The power of a test can be calculated from the number of successes and failures at step (c), which tests whether or not the simulated differences between conditions were significant. Once a predefined effect size is provided, the SIMR generates new values through Monte Carlo simulations, which are able to generate random values of categorical data assuming complex probability distributions (such as binomial). Next, it produces a chart showing the power of the test for several combinations of sample sizes and effect sizes (Figure 12). Power is calculated based on the categorical response variable of 3D perception.

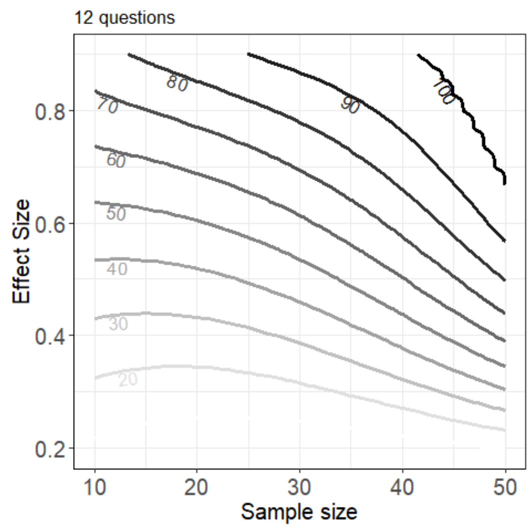


Figure 12. Power curves for main effects

In order to run these simulations, one must initially define an effect size of interest (Green and MacLeod, 2016). The effect size is entered in the SIMR tool as the magnitude of the difference expected to exist in the population. In this study, effect size is positively related to the difference in 3D perception between conditions (see item 5.1). An estimate of the difference between groups in the population is usually done through research literature, pilot study, expert judgment, and educated guessing (Gheisari, 2013). In this study, however, it was deliberately chosen to ensure practical significance of results. This initial percentage is computed into a statistics named odds ratio (OR; Szumilas, 2010) using the odds ratio formula, and then the effect size is given by log(OR) (Table 8). Although small-sized effects may exist, these are not relevant to the ultimate purpose of this study – to verify the existence of effects of immersive visualization that would benefit design review. Differences in 3D perception between conditions are only deemed beneficial/meaningful when considerably high. Therefore, the effect size was computed from an expressive difference of 33% in accuracy scores between conditions (approximately 4 hits of difference out of 12 questions, since each question accounts for approximately 8% of difference), meaning an effect size of approximately 0.69 (OR 4.88) as shown in Table 8 (scenario 4). In a within-subject study, Ozcelik and Becerik (2018) adopted an effect size of 0.7.

Table 8. Effect size calculation

Scenario	Condition		Difference	Odds Ratio (OR)	Effect Size (log(OR))
	IVR	niVR			
1	0.58	0.5	0.08 (1 hit/12 q)	1.38	0.14
2	0.67	0.5	0.17 (2 hits/12 q)	2.03	0.31
3	0.75	0.5	0.25 (3 hits/12 q)	3.00	0.48
4	0.83	0.5	0.33 (4 hits/12 q)	4.88	0.69

The calculated effect size yields a sample size of approximately 38 participants on the 80% power curve provided by the SIMR R package power analysis (Figure 12). There is no formal standard for the statistical power but 80% is a widely accepted power value in the behavioral and social sciences (Thompson and Campbell, 2004; Green and MacLeod, 2016). In sum, a sample of 38 participants has about 80% power to detect existing differences of 33% and up.

4.2.7 Recruitment of Participants

The researcher recruited 38 people according to the following qualification/inclusion criteria: experience in design review (including both classroom and industry experience), within 18-69 years of age, minimum educational level equal or over completed high school. The study population exclusion criteria are: no experience in design review, people with less than 18 or over 69 years of age, no complete high school education, non-English speakers (Lazar et al., 2017). The population of interest consists of the body of industry workforce (current and future workforce) traditionally involved in collaborative design review.

An initial screening of graduate students in the schools of Building Construction and Architecture at Georgia Tech aimed at identifying potential participants according to the qualification criteria. They were informed about the study during classes or through informal conversations (word-of-mouth). Next, a formal invitation with general information about the experiment was emailed to those who were identified as potential participants. Members of professional organizations (AGC and ABC) and of Georgia Tech's Design & Construction department were also invited by email. The volunteers who accepted the invitation to take part in the study were asked to select a time window in an online spreadsheet and come to the experiment location at the agreed time (Caddell Building on the Georgia Tech campus).

4.2.8 Assignment of Participants to Experimental Conditions

Kuliga et al. (2015) highlight the possibility of a relationship between the effects of different virtual environments and the sequence in which participants are exposed to these different simulations. In another study, Ziemer et al. (2009) examined how the order in which people experience real and virtual environments influences their distance judgments.

Therefore, in within-group experiments, random and balanced assignment of participants to experimental conditions must be conducted. Participants must be randomly assigned to all possible sequences of conditions (randomization) in a way that each sequence is administrated to the same number of participants (counter balancing). This is done to control for order effects and address the problem of systematic similarities across successive conditions, which is the cause of practice/learning and fatigue effects (Lazar et al., 2017).

In this study, there are two possible sequences of conditions: sequence 1 and sequence 2 (Table 9). Participants are randomly assigned to these sequences using the second generator in the online randomization software at randomization.com. Selecting the "balanced permutations" option in the software ensures that half the sample is assigned to sequence 1 and the other half to sequence 2. Table 10 shows the outcome of random and balanced assignment of 38 participants to the two sequences of conditions. Data analysis will also check for any order effects, that is, any significant differences in the dependent variables between groups that received treatments in different order (group exposed to sequence 1 *vs.* group exposed to sequence 2).

Table 9. Sequences of conditions

	Order of administration	
	1st	**2nd**
Sequence 1	niVR	IVR
Sequence 2	IVR	niVR

Table 10. Distribution of participants per sequence of conditions

Sequence 1	1*	2	4	5	7	8	13	14	15	16	19	21	25	27	28	31	32	33	38
Sequence 2	3	6	9	10	11	12	17	18	20	22	23	24	26	29	30	34	35	36	37

* 1 = Participant # 1, and so on.

An additional strategy to mitigate learning effects consists in randomly sorting the order of survey questions across environments as proposed by Paes et al. (2017). Besides conducting randomization, counter balancing, and sorting the order of questions, allowing regular breaks during the experiment, as well as sufficient acquaintance time and training so that participants can get used to navigation in the virtual environments can also help to reduce the impact of learning and fatigue effects over consecutive conditions.

4.2.9 Data Collection

Data collection takes place in the experiment sessions, which comprise the fourteen steps listed below. Each session took approximately 70 minutes, in a total of 44 hours of data collection over the Spring and Fall semesters of 2019. Figure 13 shows participants performing perception tasks during experiment sessions.

1. Check equipment and instruments before participant arrives at the experiment location;

2. Greet and introduce the experiment purposes and procedures to participant;
3. Collect the participant's consent (Consent Form);
4. Collect the participant's demographic information (DQ);
5. Collect the participant's spatial ability (Spatial Ability Test);
6. Assign participant to starting VR mode, according to predefined sequence of presentation;
7. Let participant get used to the first VR system assigned (niVR or IVR mode);
8. Guide participant through 3D perception tasks in the first VR mode while administering the 3D perception questionnaire (3DPQ);
9. Administer the presence questionnaire (PQ) after walkthrough in the first VR mode;
10. Let participant get used to the second VR system assigned (niVR or IVR mode);
11. Guide participant through 3D perception tasks in the second VR mode while administering the 3D perception questionnaire (3DPQ);
12. Administer the presence questionnaire (PQ) after walkthrough in the second VR mode;
13. Guide participant through 3D perception tasks in the physical environment (PhE) while administering the 3D perception questionnaire (3DPQ);
14. Briefing and thank participant.

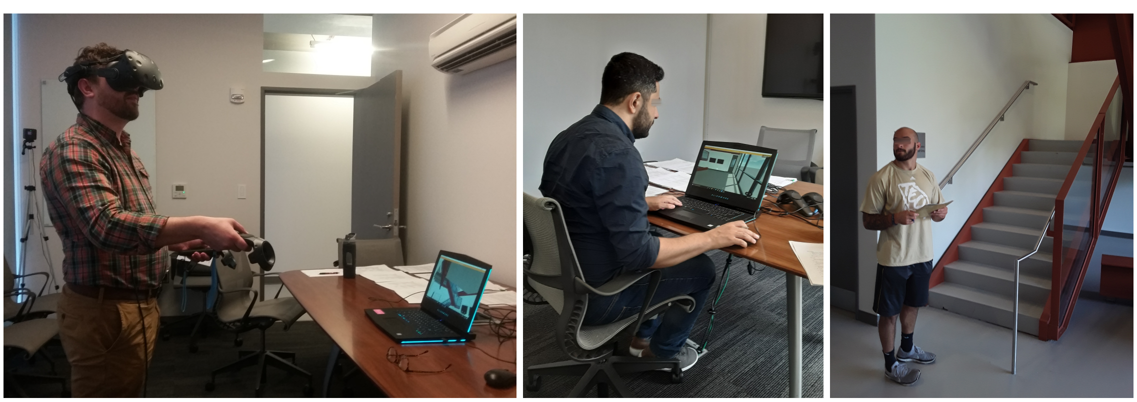


Figure 13. Participants performing perception tasks: in IVR mode (left), in niVR mode (center), and at the physical environment (right)

CHAPTER 5

DATA ANALYSIS & RESULTS

The goal of data analysis is to provide a set of descriptive and inferential statistics that describe experimental data and infer the relationships between variables and how they impact each other. Preprocessing of data for statistical analysis involved cleaning up, coding, and organizing data for specific statistical software (PAST software application – Hammer, 2001).

Descriptive statistics are used to understand the nature of the data set. This study utilizes percentages, means (arithmetic averages) and standard deviations (indicate how the data set is distributed) as measures of spread, medians (middle score in the data set), and modes (most frequent value in the data set) when appropriate. In turn, inferential statistics are used to examine: a) whether differences in 3D perception and presence responses between experimental conditions, among questions, and between conditions per question are significant (H1 and H2), b) whether 3D perception and presence responses are associated within conditions (H3), and c) the influence of individual factors (confounding variables) and sequence/order of presentation of conditions on 3D perception and presence responses. Hypothesis testing is conducted through significance tests, which provide the statistical significance of the relationships observed. The tests estimate the probability of differences assuming that the null hypothesis is true (p value), establishing the level of confidence with which the relationships observed can be generalized to the entire population.

Various significance tests can be used to test statistical hypotheses. The first step in determining which significance tests are appropriate is to examine the nature of the data set and assign it a reasonable probability distribution (Lazar et al., 2017). Real data follow no probability distribution so that the assignment of probability distributions to field-collected data will always represent an approximation. There are several probability distributions and specific tests for each one. Parametric significance tests commonly used in the social and behavioral sciences such as the t test and the Analysis of Variance (ANOVA) assume that the data set is approximately normally distributed (i.e., it follows a normal/Gaussian distribution), but this is rarely true (Rosenthal and Rosnow, 2008). It is also rarely true that whether normal distribution cannot be assumed, nonparametric tests should be necessarily adopted. There are other parametric methods within the group of Generalized Linear Models (GLM) that are not constrained by the normality assumption, such as regression models. Besides, nonparametric tests are not free of limitations. Because they convert the original data into ranks, information can be lost when the data tested are actually interval or ratio so all that matters is the rank/order of the data while the distance information between data points is lost (Lazar et al., 2017). This is the reason why these tests are more adequate for treating categorical ordinal data (distributed into ordered categories, such as "little, moderate, great extent"), which are inherently poor in distance information. In addition, the most common nonparametric tests – Wilcoxon Signed-Rank, Mann-Whitney U, Friedman's, and Kruskal-Wallis – can only be used to analyze data that involves only one independent variable. If the data are not normally distributed, one may consider to: a) transform the data so they are normally distributed, b) adopt nonparametric tests for the analysis (Lazar et al., 2017), or c) assign other probability distributions to the data set such as binomial, multinomial, Poisson, and gamma distributions, for which there are specific testing methods, including regression models.

The literature indicates that when analyses involve non-normal variables, either nonparametric tests or regression models may be used. In this study, however, both nonparametric and parametric tests (based on the normality assumption) were deemed limited for data analysis. Parametric tests were deemed unfitted due to two main reasons: a) one of the dependent variables is not normally distributed (3D perception), and b) would require transforming/collapsing categorical nominal data (3D perception) into percentages, leading to information loss. In turn, nonparametric tests would work better for categorical ordinal data, as they convert data into ranks. Also, generally, they can support only one independent variable. Therefore, this study utilizes several regression models to conduct hypothesis testing.

Data analysis utilizes a particular class of semiparametric regression models known as Generalized Estimating Equation (GEE; Liang and Zeger, 1986) – a robust model deemed more effective than analyses of variance for repeated measures designs. The GEE is a marginal regression model often used in longitudinal/clustered data analysis in which there are repeated measures of the same subject. It consists of a very flexible class of models, capable of accommodating several probability distributions of the response variable (e.g., binomial, gamma, normal, Poisson) and many covariance structures (Wang, 2014). The model computes the mean values of the dependent variable in the groups and then provides the significance of the difference between those means through the *p* value of the line slope. Thus, through a regression analysis it is possible to determine, for instance, whether an experimental condition affects user 3D perception and presence (Rosenthal and Rosnow, 2008).

P values are provided throughout data analysis to indicate the significance of differences and relationships observed. Also known as significance value, the *p* value is the probability of differences assuming that the null hypothesis is true. When such probability is very low, below a pre-defined threshold value (significance level), it allows an investigator to reject the null hypothesis. According to past studies in the field, this research adopts a significance level (α) of 0.05, hence a confidence level of 95% ($1 - \alpha$). Whether a *p* value is below that significance level of 0.05, the null hypothesis can be correctly rejected, and one may claim with 95% of confidence that the differences observed are statistically significant and due to differences in the controlled independent variables.

Past Similar Analyses

In a thorough study conducted by Bafna and Chambers (2014), regression models were used to check the relationship between the spatial organization of an apartment (predictor variable) and the levels of inhabitant activities around it (dependent variable). The predictor variable was treated as continuous data ("interconnectedness"), whereas the dependent variable was treated as count data (hours) and modeled assuming Poison distribution. The researchers used multiple regression analyses to control for the effects of age and educational level on the dependent variable (hours), and found that these did not affect the main relationship. They entered the confounding variables separately first and then at the same time into the model, and spatial organization remained a significant factor in the explanation of activity hours. In a similar study, Chambers et al. (2018) examined the relationship between apartment layout (predictor variable) and the odds of depression (response variable) controlling for demographics, health conditions, and housing and neighborhood characteristics. They utilized ANOVA and Chi-squared test to determine the relationship between both values of the predictor variable and gender with depression symptoms. In addition, logistic regression models were adjusted to examine the relationship between apartment layout and depression symptoms per gender controlling for covariates.

In order to test the effects of conditions on presence, Khashe et al. (2018) utilized independent samples t test, whereas to verify interactions between dependent variables (presence and task performance) within each condition they adjusted regression models. Castronovo et al. (2017) utilized the nonparametric Wilcoxon Signed-Rank test for paired samples to compare presence levels between conditions, and the nonparametric Mann-Whitney U test for independent samples to check for order effects (entering order of presentation as a between-subject factor). Kinsey et al. (2006) utilized t test to determine the significance of differences in spatial ability between groups of participants, as well as the effect of gender (entered as the independent variable) on spatial ability (although they have not sampled gender as an independent variable). Then, they utilized Pearson's correlation coefficient test to verify the association between the two dependent variables (spatial ability and self-efficacy).

Ozcelik and Becerik (2018) conducted a within-subject study to validate VEs as adequate simulations of physical environments by comparing users' perceived thermal comfort between virtual- and physical-world conditions. Because data were not normally distributed they decided to adopt the nonparametric Wilcoxon Signed-Rank test (paired samples) to analyze the differences in perceived thermal comfort (hit/fail categorical variable) between conditions, and the Mann-Whitney U test (independent samples) to check for differences in age and gender between conditions. In this case they possibly ran two separate tests: one for checking the "effects" of conditions (independent variable) on age (dependent categorical variable), and another one to check the "effects" of conditions on gender. To investigate the association between two dependent variables (presence and individual factor) they also utilized Pearson's test.

In a between-subject study conducted by Kimura et al. (2017), in which they compare participants' orientation performance between virtual- and physical-world conditions, binomial tests were used to examine the ratio of successes over fails in selecting the correct target within each condition separately. Next, they employed independent samples *z* test to check the significance of the difference in the percentage of success between conditions. Brookes et al. (2019) compared the effects of two virtual environments (static *vs.* oscillating) on participants' postural sway. They designed a split-plot experiment and utilized a mixed-model ANOVA to test the significance of differences observed. In a between-subject study, Higuera-Trujillo et al. (2017) utilized Mann-Whitney U test (independent samples) to examine the differences in psychological and physiological responses between a display format and the physical environment. They ran one test for each dependent variable in each of their three conditions comprising a display format and the physical environment (disp1-phe; disp2-phe; disp3-phe).

They also ran correlation analyses to examine possible associations between the psychological and the physiological responses.

Regression Analysis

Regression analyses are used for two complementary purposes: hypothesis testing and predictions (Lazar et al., 2017). The aim of adjusting a regression model for hypothesis testing is to examine the relationships between one dependent variable and one or more independent variables. In other words, the goal is to adjust a model/equation based on the independent variables that best explains the variances in the dependent variable, so that the statistical hypothesis about the fitted coefficients of the equation can be tested. This fitted model can eventually be used to make predictions. That is, in order for a regression model to make valid and useful predictions, it must be significant. In this case, values of the independent variables (also referred to as predictor variables) can be used to predict/estimate the value of the dependent variable (Pujoni, 2019). The two aforementioned aims are closely related. However, in this study, regression models are used only for hypothesis testing and not for making predictions.

A linear regression analysis plots a regression line (with linear and angular coefficients) utilized to estimate the means of data points. The regression line is where these mean values lie, providing the mathematical equation/model to predict means of new data points. Whether the slope of a regression line is significantly different from zero (provided by the p value of its slope/angular coefficient), the means are significantly different from each other, meaning that the independent variable has a significant effect on the mean of the dependent variable. Therefore, the regression model is the test of a statistical hypothesis itself. Consequently – and if needed – that equation can be utilized to predict the mean values of the dependent variable given new values of the predictor variable. Ultimately, predictions can only be made if the hypothesis testing

– the analysis of the significance of the relationship between predictor and dependent variables – produces a slope value significantly different from zero.

Parametric statistical tests and regression models are quite similar methods – they all pertain to the group of Generalized Linear Models (GLM) – that estimate the mean values of each group (in ANOVA) or the mean values of individual data points as if each point were its own "group" (in linear regression), followed by an analysis of variance that divides the variance of group means in relation to the global variance (explained variance) by the variance within groups (residual variance). ANOVA, t test, and regression models are all able to test the significance of differences in the dependent variable in relation to changes in the value of the predictor variable(s).

If a categorical nominal variable with two levels (such as conditions "niVR" and "IVR") is entered on the X-axis of a regression model (the axis for predictor variables), and the slope of the line connecting the means of the dependent variable in the groups is significantly different from zero, the means are significantly different from each other. In this situation, adjusting a regression model corresponds to testing the significance of the difference between the group means. If the dependent variable were continuous, the adjusted model would be a simple ANOVA.

A correlation is also similar to linear regression in the sense that both can provide the linear association between two variables. A correlation is half way to a regression model. The correlation coefficient (r) is simply the square root of the coefficient of determination (r^2, or "R-squared") of a simple linear regression. Most statistical software applications generate both statistics after adjusting a regression model. There are, however, a few critical differences. First, while correlations allow the study of the association between two variables only, multiple regression models can check the relationship between one dependent variable and a number of independent/predictor variables, computing separately the correlations between each predictor with the dependent variable and testing the significance of each association individually. Second, as opposed to regression models, the Pearson's correlation coefficient test (the parametric correlation method) requires meeting the assumptions of normality of both variables (predictor and dependent) and linearity of the relationship (an assumption that the relationship is linear). Third, as opposed to correlations, regression models can be used to make predictions in addition to providing the association significance. Lastly, while variables in a correlation analysis are interchangeable, that is, the correlation between X and Y variables is the same as the correlation between Y and X, in a simple regression these variables are fixed, that is, a regression of Y on X differs from a regression of X on Y (Pujoni, 2019).

A variable entered in the X-axis of a regression model is considered an independent variable even if it was not actually controlled by the investigator. When there is only one independent variable, the number of random (dependent) variables determines whether it is called a multivariate or univariate model. A univariate model has a single dependent variable (simple linear regression), whereas a multivariate model can have two or more dependent variables. When there are two or more independent variables predicting a single response variable, this is called a multiple model. Multiple regression models are considered an adequate and flexible statistical method to account/control for potentially confounding effects. They can handle multiple confounding variables simultaneously as predictor variables (or covariates), allowing for the study of the relationships between these and the response variable (Pourhoseingholi et al., 2012).

Sample Demographics

The sample size is of 38 participants – as defined by *a priori* power analysis – and complies with the qualification/inclusion criteria: experience in design review (including both classroom and industry experience), within 18-69 years of age, minimum educational level equal or over completed high school. The study's population of interest consists of the body of industry's current and future workforce traditionally involved in collaborative design review.

Considering the entire sample, most participants are attending graduate school (94.7%) and 42.1% of them have between 26 and 33 years of age. Gender distribution in the sample is relatively and unintentionally balanced (female = 44.7%; male = 55.3%). In terms of their Bachelor's degree majors, 28.9% are Architects, 39.5% are Civil Engineers, and 5.3% are both. When it comes to their current academic majors, approximately half of the sample comprises Building Construction students (52.6%), 13.2% are Architecture students, and 13.2% are Civil Engineering students. The participants have taken different professional roles during their industry experiences – oftentimes more than one occupation: 36.7% have worked as Architects, 39.4% as Civil Engineers, and 18.3% as Construction Managers. Only four participants (10.5%) do not have any industry experience (although they do have classroom experience in design review). Their experience in design review is relatively balanced across the first three ranges of experience: 21.1% have up to 1 year, 26.3% have between 1 and 5 years, 23.7% have between 5 and 10 years, and 29% have over 10 years. All participants reported regular computer usage. Over half the participants reported to have either beginner (28.9%) or intermediate (23.7%) level of experience with 3D virtual environments (3D modeling and BIM software, videogames, etc.), whereas 39.5% stated to have expert level of experience. In regards to their familiarity with the experiment environment (the Caddell Building lobby), most participants stated to pay regular visits to the location (57.9%), 26.3% are there occasionally, and 13.2% have never been to the lobby. All participants scored between 6 and 10 in the Spatial Ability Test (out of 10 questions). Over half the participants scored between 7 and 9 in a relatively balanced distribution, as follows: 15.8% scored 7 points, 21.1% scored 8 points, 18.4% scored 9 points. Interestingly, 39.5% of the participants achieved a 10 score. Table 11 summarizes the sample demographics.

Table 11. Sample demographics

Parameter	Entire sample (n = 38)		Sequence 1: niVR – IVR (n = 19)		Sequence 2: IVR – niVR (n = 19)	
	#	%	#	%	#	%
Age						
18 – 25	10	26.3%	6	31.6%	4	21.1%
26 – 33	16	42.1%	5	26.3%	11	57.9%
34 – 41	7	18.4%	3	15.8%	4	21.1%
50 – 69	5	13.2%	5	26.3%	0	-
Gender						
Female	17	44.7%	4	21.1%	13	68.4%
Male	21	55.3%	15	78.9%	6	31.6%
Educational level (completed or ongoing)						
Bachelor's degree	2	5.3%	2	10.5%	0	-
Master's degree	25	65.8%	12	63.2%	13	68.4%
Doctoral degree	11	28.9%	5	26.3%	6	31.6%
Bachelor's major						
Architecture	11	28.9%	5	26.3%	6	31.6%
Architecture, Civil Engineering	2	5.3%	2	10.5%	0	-
Building Environment	1	2.6%	1	5.3%	0	-
Civil Engineering	15	39.5%	5	26.3%	10	52.6%
Environmental Science and Engineering	3	7.9%	2	10.5%	1	5.3%
Landscape Architecture	1	2.6%	1	5.3%	0	-
Other	5	13.2%	3	15.8%	2	10.5%
Current academic major						
Architecture	5	13.2%	3	15.8%	2	10.5%
Building Construction	20	52.6%	9	47.4%	11	57.9%
Civil Engineering	5	13.2%	2	10.5%	3	15.8%
Environmental Engineering	1	2.6%	0	-	1	5.3%
Industrial Design, Building Construction	1	2.6%	1	5.3%	0	-
N/A – Not enrolled	6	15.8%	4	21.1%	2	10.5%
Professional Occupation(s)						
Architect	7	18.4%	3	15.8%	4	21.1%
Architect, Civil Engineer	1	2.6%	1	5.3%	0	-
Architect, Civil Engineer, Faculty	1	2.6%	1	5.3%	0	-
Architect, Construction Mngr., Facility Mngr.	1	2.6%	1	5.3%	0	-
Architect, Construction Mngr., Faculty	1	2.6%	1	5.3%	0	-
Architect, Construction Mngr., Trade Contractor	1	2.6%	1	5.3%	0	-
Architect, Faculty	2	5.3%	0	-	2	10.5%
Civil Engineer	10	26.3%	2	10.5%	8	42.1%
Civil Engineer, Construction Mngr.	1	2.6%	0	-	1	5.3%
Civil Engineer, Faculty	2	5.3%	1	5.3%	1	5.3%
Construction Manager	2	5.3%	1	5.3%	1	5.3%
Environmental Engineer	2	5.3%	0	-	2	10.5%
Environmental Engineer, Construction Mngr.	1	2.6%	1	5.3%	0	-
Facility Manager	1	2.6%	1	5.3%	0	-

Landscape Architect	1	2.6%	1	5.3%	0	-
Student (only)	4	10.5%	4	21.1%	0	-
Experience in design review						
Up to 1 year	8	21.1%	4	21.1%	4	21.1%
1 – 5 years	10	26.3%	4	21.1%	6	31.6%
5 – 10 years	9	23.7%	3	15.8%	6	31.6%
10 – 15 years	5	13.2%	2	10.5%	3	15.8%
15 – 20 years	2	5.3%	2	10.5%	0	-
20 + years	4	10.5%	4	21.1%	0	-
Computer usage						
Regular	38	100%	19	100%	19	100%
Experience with 3D virtual environments						
None	3	7.9%	1	5.3%	2	10.5%
Beginner	11	28.9%	5	26.3%	6	31.6%
Intermediate	9	23.7%	5	26.3%	4	21.1%
Expert	15	39.5%	8	42.1%	7	36.8%
Familiarity with the experiment environment						
None	5	13.2%	2	10.5%	3	15.8%
Rare visits	1	2.6%	1	5.3%	0	-
Occasional visits	10	26.3%	5	26.3%	5	26.3%
Regular visits	22	57.9%	11	57.9%	11	57.9%
Spatial ability score (out of 10)						
6 score	2	5.3%	2	10.5%	0	-
7 score	6	15.8%	3	15.8%	3	15.8%
8 score	8	21.1%	2	10.5%	6	31.6%
9 score	7	18.4%	5	26.3%	2	10.5%
10 score	15	39.5%	7	36.8%	8	42.1%

An association analysis is conducted to verify which individual factors (confounding variables) are associated with each other within the study's sample. It is important to know which factors are associated to ensure accuracy and validly of the analysis of influence of such factors on the dependent variables, which can be found in item 5.4. Naturally, the effects of factors that are associated cannot be interpreted separately. For instance, if educational level is associated with age, and it affects a dependent variable, it would not be possible to attribute that effect to educational level only, as it could also have been due to age. The individual factors examined are nine: age, gender, educational level, bachelor's major, current major, experience in design review, experience with 3D virtual environments, familiarity with the experiment environment, and spatial ability. Computer usage was not analyzed since all participants reported the same usage level (regular). A participant's bachelor's major will be interpreted as one's main professional role or original field of practice. Because confounding variables are categorical, evaluating association among them involves developing contingency tables of categorical data, also known as tables of counts, and computing *p* values using the nonparametric Chi-squared test of independence, which is adequate for treating categorical nominal data from independent samples (Lazar et al., 2017). The significant associations found are restricted to this study's sample. Table 12 provides an example of a contingency table utilized for this analysis.

Table 12. Contingency table utilized for the analysis of association among individual factors

Exp. in design review / **Age**	**Up to 1**	**1-5**	**5-10**	**10-15**	**15-20**	**20+**
18-25	4*	4	2	0	0	0
26-33	4	4	7	1	0	0
34-41	0	2	0	4	1	0
50+	0	0	0	0	1	4
Chi p value	<0.001					
Association strength	0.7					

* # of participants

A *p* value below 0.05 indicates a significant association between two confounders. The association strength is given by the association strength index. Values within 0-0.3 are weak associations, medium strength associations are within 0.3-0.7, and values over 0.7 are strong associations. Table 13 shows the significant associations among individual factors in the sample.

Table 13. Associations among individual factors

Factor 1	**Factor 2**	**P value**	**Association Strength**
Age	Bachelor's major	0.338	0.535
	Current major	0.0501	0.468
	Educational level	**0.00454**	**0.497**
	Experience in design review	**< 0.001**	**0.7**
	Experience with 3D VEs	0.611	0.252
	Familiarity with the exp. env.	0.716	0.234
	Gender	0.103	0.404
	Spatial ability	0.0999	0.403
Bachelor's major	Current major	**0.0021**	**0.663**
	Educational level	**0.0272**	**0.667**
	Experience in design review	0.133	0.568
	Experience with 3D VEs	0.693	0.474
	Familiarity with the exp. env.	0.206	0.562

Bachelor's major	Gender	0.449	0.511
	Spatial ability	0.492	0.51
Current major	Educational level	**<0.001**	**0.646**
	Experience in design review	0.316	0.383
	Experience with 3D VEs	0.487	0.357
	Familiarity with the exp. env.	**<0.001**	**0.672**
	Gender	0.575	0.317
	Spatial ability	0.523	0.353
Educational level	Experience in design review	0.0712	0.475
	Experience with 3D VEs	**0.0365**	**0.421**
	Familiarity with the exp. env.	**0.0137**	**0.459**
	Gender	0.421	0.213
	Spatial ability	0.787	0.249
Experience in design review	Experience with 3D VEs	0.798	0.301
	Familiarity with the exp. env.	0.143	0.427
	Gender	0.304	0.398
	Spatial ability	0.224	0.401
Experience with 3D VEs	Familiarity with the exp. env.	**0.0367**	**0.396**
	Gender	0.655	0.207
	Spatial ability	**0.0257**	**0.452**
Familiarity with the exp. env.	Gender	0.315	0.305
	Spatial ability	0.464	0.321
Gender	Spatial ability	0.583	0.274

The analysis of association among individual factors revealed ten significant associations among eight factors (gender – the ninth factor – was not associated with any other factor). Figure 14 shows those eight factors and their associations represented by lines. In green are the factors that will be tested for their impacts on the dependent variables (see item 5.4). Given this study's sampling process and sample characteristics, some significant associations were expected, for instance, between age and educational level, age and experience in design review, and current major and familiarity with the experiment environment.

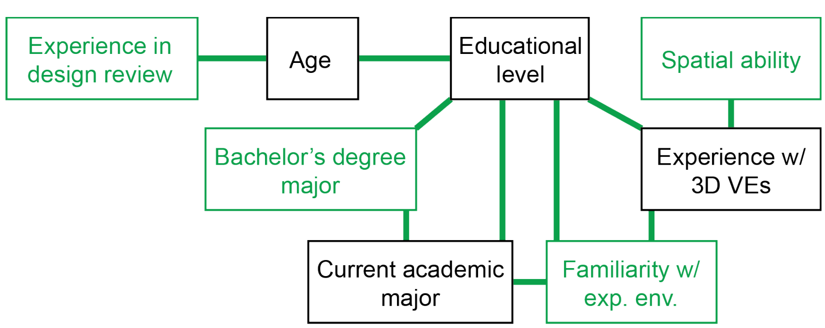


Figure 14. Significant associations among individual factors

It should be noted that individual factors were not deliberately balanced between the groups of sequence of presentation, which resulted in expected discrepancies of demographic characteristics between those groups as shown in Table 13. Gender is not balanced between the groups exposed to sequences 1 and 2 – i.e., there is a significant association between gender and sequence (p = 0.009). While sequence 1 encompasses a larger number of male participants (15 male, 4 female), sequence 2 comprises a larger number of female participants (6 male, 13 female). Bachelor's major is not balanced as well (a nearly significant association, p = 0.0505). The number of civil engineers in the group exposed to sequence 1 is half (5) the number in the group of sequence 2 (10). Table 14 provides the associations between sequence of presentation and individual factors in the sample.

Table 14. Associations between sequence of presentation and individual factors

	Individual factor	**P value**	**Association Strength**
Sequence of presentation	Age	0.0505	0.453
	Bachelor's major	0.35	0.54
	Current major	0.659	0.293
	Educational level	0.345	0.237
	Experience in design review	0.18	0.447
	Experience with 3D VEs	0.896	0.126
	Familiarity with the exp. env.	0.753	0.178
	Gender	**0.009**	**0.476**
	Spatial ability	0.253	0.375

5.1 H1 – Analysis of 3D Perception

In regards to participants' 3D perception performance within the VR modes, it was not expected that they would be accurate in perceiving actual distances from the environment depicted. As previously discussed, experiments of this nature should not expect that participants are able to correctly estimate, for instance, the actual ceiling height. Again, the initial goal is to compare a participant's 3D perception when visiting the physical environment (reference mode) with her/his perceptions in the virtual environments. Therefore, the answer to a question of the 3DPQ administered in each VR mode (3DPQ-niVR and 3DPQ-IVR) is compared to the response to the very same question administered in the physical environment (3DPQ-PhE). The compatibility between these responses indicates the ability of a virtual environment to reproduce 3D perception obtained in the physical environment, hereafter referred to as accuracy: the degree of resemblance to 3D perception in the real world. In other words, accuracy refers to the resulting deviation of a participant's 3D perception in a virtual environment with respect to her/his perception in the real world. The *accuracy score* is given by the ratio of hits over the total number of observations. A hit is defined as when a participant selects the same alternative option to a given question of the 3DPQ in both reference mode and VR mode. Ziemer et al. (2009) also adopted the term accuracy score. Higuera-Trujillo et al. (2017) developed similar methodology, naming closeness what in this study is called accuracy. Their closeness scores were also obtained from the "standardization" of virtual-world responses over physical-world responses to simplify the comparisons among display formats.

Code numbers are assigned to hit and error categories: 1 represents a hit, and 0 represents an error/fail. Ozcelik and Becerik (2018) conducted a similar study in which the first step was to compare participants' perceived temperature to the actual room temperature. They adopted 1 and 0 values to represent when a participant correctly guessed the actual temperature (within an error margin) or not, respectively. Kimura et al. (2017) also utilized 1 and 0 to represent a correct response and an incorrect response, respectively. In this study, 1 and 0 values are derived from comparisons of participants' 3D perception between a virtual environment and the real environment, not between real environment and actual dimensions. Ultimately, these comparisons will generate accuracy scores (per participant, per question, per condition), which will serve for all subsequent analyses (Paes et al., 2017). An example of 3D perception accuracy score calculation per participant is provided in Table 15.

Table 15. Example of accuracy score calculation per participant

3DPQ question	**Chosen Alternative Option**		**Hit**	**Error**
	VR mode (niVR or IVR)	**Reference mode** (PhE)		
1	B	C		0
2	F	D		0
3	A	A	1	
12	…	…	…	…
		# of Hits	5	
		# of Errors		7
		Accuracy Score	42%	

Accuracy scores per VR modes are compared to determine which condition better reproduced a participant's 3D perception in the real world, that is, which technology "did better" in reproducing a participant's 3D perception. The condition where participants achieve significantly higher accuracy scores provides the most similar 3D perception to that obtained in the real world, and it could be deemed the technology that offers the most realistic 3D perception experience of built spaces.

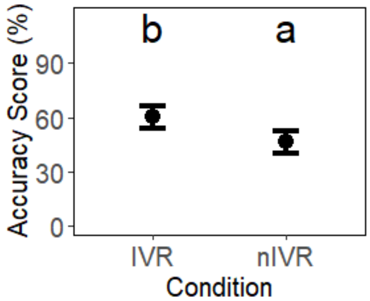
5.1.1 Significance Test

The accuracy scores data set consists of categorical nominal values (hit/error values coded into one/zero values). This data set assumes a binomial probability distribution, which is adequate for treating categorical (nominal) responses. The total number of hit/error responses should have been 456 (38 participants *x* 12 questions) per condition. However, three observations were discarded due to the participant's inability to provide an answer, totalizing 453 values.

Three main relationships were tested: a) the difference of global accuracy scores between conditions, b) the difference of global accuracy scores across questions, and c) the difference of accuracy scores between conditions, per question and group of questions. The estimated parameters of the adjusted models were extracted along with their respective 95% confidence interval defined by lower and upper confidence boundaries. The confidence interval is the range of values within which the population parameters lie (mean accuracy scores) with 95% of confidence. Standard error values – a measure of distance of all possible sample means from the population mean – are provided as well (Pujoni, 2019). The letters above each group in the following charts represent the comparisons. Groups that do not share the same letters are significantly different from each other (0.05 of significance). In the analysis per questions, Tukey's correction was applied on the p values to avoid inflation of type I error due to multiple comparisons.

5.1.1.1 Difference of global accuracy scores between conditions

Conditions were included as predictor variables of accuracy scores and the significance of the difference of global accuracy scores between conditions was tested (simple linear regression). Figure 15 provides a chart showing the global accuracy scores in each VR mode, along with a table with detailed information on the scores. The global accuracy scores differ between IVR and niVR conditions ($p < 0.001$), with higher values found in the IVR condition. Therefore, we refute the null hypothesis of no difference in 3D perception between VR modes.



Condition	niVR	IVR
Hit/Error Occurrence	211/242	273/180
Accuracy Score	47%	60%
Standard Error	3%	3%
Lower confidence limit	40%	54%
Upper confidence limit	53%	66%
Group	a	b

Figure 15. Global accuracy scores in the VR modes

5.1.1.2 Difference of global accuracy scores among questions

Differences of global accuracy scores among questions were also tested. This helps in the determination of the efficacy and validity of the 3DPQ questionnaire. This analysis encompasses both conditions and focuses on evaluating how each question did in measuring the dependent variable. In this case, 3DPQ questions were included as predictor variables of accuracy scores (simple linear regression). Figure 16 shows the global accuracy scores per question. Table 16 provides detailed information on the global accuracy scores for each question. Pairwise comparisons (all possible pairs of questions) show that the global accuracy scores are different within three pairs of questions. The global accuracy score of Q12 differs from the global accuracy scores of Q3 (p = 0.017) and Q4 (p = 0.001), with lower values found in Q12. The global accuracy score of Q4 also differs from the global accuracy score of Q8 (p = 0.015), with lower values in Q8. Questions 1, 2, 5, 6, 7, 9, 10 and 11 do not differ from any other question – neither from Q3 and Q4, which yielded the highest accuracy scores, nor from Q12, which yielded the lowest score.

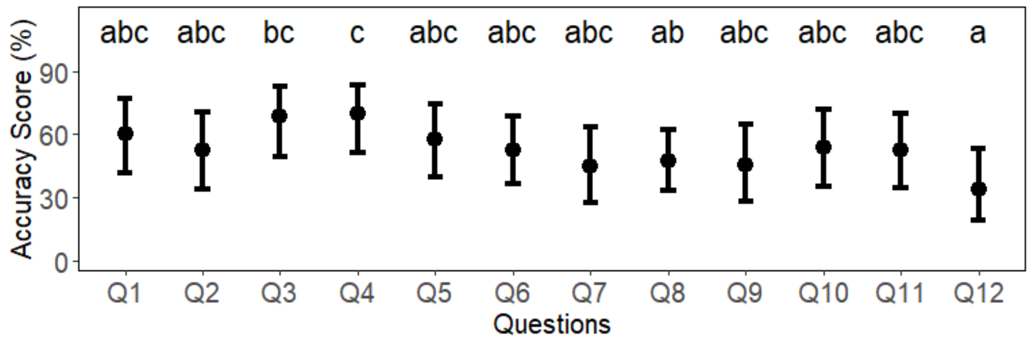


Figure 16. Global accuracy scores per question

Table 16. Global accuracy scores per question

Questions	Accuracy Score	Standard Error	Lower confidence limit	Upper confidence limit	Group
Q1	61%	6%	41%	77%	abc
Q2	53%	7%	34%	71%	abc
Q3	68%	6%	49%	83%	bc
Q4	70%	6%	51%	83%	c
Q5	58%	6%	40%	74%	abc
Q6	53%	6%	36%	69%	abc
Q7	45%	7%	27%	64%	abc
Q8	47%	5%	33%	62%	ab
Q9	46%	7%	28%	65%	abc
Q10	54%	7%	35%	72%	abc
Q11	52%	7%	34%	70%	abc
Q12	34%	6%	19%	53%	a

The results indicate consistency in the level of difficulty across 3DPQ questions. The researcher conducting the experiment sessions noticed that in questions 3 and 4 – the ones that yielded the highest accuracy scores – participants strongly relied on their knowledge about elements in the space, which possibly ended acting as major sources of depth cues (the staircase steps in question 3, and the lobby's entrance door in question 4). It appears that the information provided by those anchor elements made participants keep with their answers between VR modes and physical environment – given those depth cues they strongly believed their answers to be right, regardless of the environment. Question 12 was expected to yield low accuracy scores due to its difficulty, since it prompts participants to estimate a particularly short dimension as opposed to the other questions – the alternative answer options in Q12 provide distance ranges within inches/centimeters.

5.1.1.3 Difference of accuracy scores between conditions, per question and group of questions

This analysis looks into the difference of accuracy scores in the interaction between conditions and questions (Figure 17). It aims at examining the extent to which each question contributed to the difference of global accuracy scores between conditions. Table 17 provides detailed information on the difference of accuracy scores between conditions for each question. Conditions, questions, and the interaction between them were included as predictor variables of accuracy scores (multiple regression model with three predictor variables). The accuracy scores differ between IVR and niVR conditions for questions Q3 (p = 0.002), Q5 (p = 0.027), and Q12 (p = 0.002), with higher values found in the IVR condition. In all questions, the IVR condition shows higher accuracy scores, whether statistically significant or not.

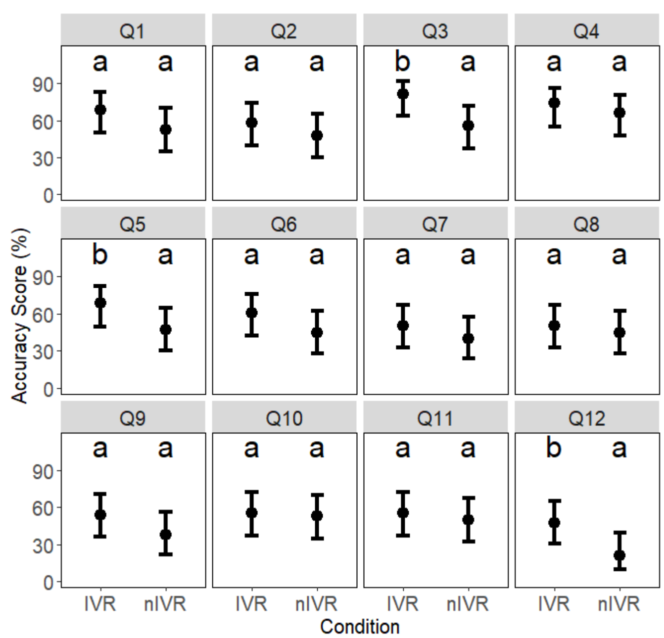


Figure 17. Accuracy scores between conditions, per question

Table 17. Accuracy scores between conditions, per question

Questions	Condition	Accuracy Score	Difference	Standard Error	Lower confidence limit	Upper confidence limit	Group	P value
Q1	niVR	53%	16%	8%	35%	70%	a	0.076
	IVR	68%		8%	50%	83%	a	
Q2	niVR	47%	11%	8%	30%	65%	a	0.243
	IVR	58%		8%	40%	74%	a	
Q3	niVR	55%	26%	8%	37%	72%	a	**0.002**
	IVR	82%		6%	63%	92%	b	
Q4	niVR	66%	8%	8%	47%	81%	a	0.404
	IVR	74%		7%	55%	86%	a	
Q5	niVR	47%	21%	8%	30%	65%	a	**0.027**
	IVR	68%		8%	50%	83%	b	
Q6	niVR	45%	16%	8%	28%	63%	a	0.153
	IVR	61%		8%	42%	76%	a	
Q7	niVR	39%	11%	8%	24%	58%	a	0.243
	IVR	50%		8%	33%	67%	a	
Q8	niVR	45%	5%	8%	28%	63%	a	0.670
	IVR	50%		8%	33%	67%	a	
Q9	niVR	38%	16%	8%	22%	56%	a	0.076
	IVR	54%		8%	36%	71%	a	
Q10	niVR	53%	3%	8%	35%	70%	a	0.763
	IVR	55%		8%	37%	72%	a	
Q11	niVR	50%	6%	8%	32%	67%	a	0.592
	IVR	55%		8%	37%	72%	a	
Q12	niVR	21%	26%	7%	10%	39%	a	**0.002**
	IVR	47%		8%	30%	65%	b	

As revealed by the previous analysis, Q12 was the most difficult question (both conditions considered), which means that participants could not estimate that dimension accurately in the VR modes in relation to their real-world perceptions. However, Q12 detected a significant difference between conditions in favor of IVR meaning that, while it was a difficult question, participants did better in it using IVR. Similarly, Q3, where participants achieved one of the highest accuracy scores considering both conditions meaning that they were able to estimate that dimension quite accurately in the VR modes, also detected a significant difference between conditions in favor of IVR. That is, while Q3 was an easy question, participants did significantly better in it using IVR.

Question 4, in which participants also did relatively well considering both conditions (in the previous analysis), could not detect a significant difference between conditions. In fact, the non-significant difference yielded by Q4 was one of the lowest among all questions (8%). The fact that while Q4 was almost as easy as Q3 but it did not detect a significant difference between conditions as Q3 did is possibly due to the different types of distance estimation they prompt. Question 3 is one of three questions of egocentric depth estimation at fixed position, whereas Q4 is one of three questions of interobject depth estimation at fixed position. That is, while Q3 gives egocentric depth estimation, Q4 addresses interobject depth estimation.

Question 5 was also capable of detecting a significant difference of accuracy scores between conditions. This question is quite similar to Q4 and it also addresses interobject depth estimation. However, in this question IVR provided a significant improvement. The fact that Q4 and Q5 measure the same thing but Q4 did not detect a significant difference between conditions whereas Q5 did reinforces the issue with Q4's statement, more specifically, with the fact that participants could rely on their knowledge about the width of the lobby's entrance door, which acted as a dominant depth cue in both conditions.

As expected, in each question separately the confidence interval is wider than the confidence interval of the global accuracy scores in the VR modes, since the variance of data within each question in much larger given the smaller data set per question (456 observations divided by 12 questions, totalizing 38 data points per question). Therefore, the probability of detecting significant differences in each question separately was much lower. Nonetheless, in Q3, Q5, and Q12 the effect of conditions were so strong that even given a critically smaller data set and larger variance, the difference between conditions was still significant. Such strong effects offset the other non-significant differences in other questions, resulting in a highly significant difference of global accuracy scores between conditions. Again, as expected, the non-significant differences found in nine questions were due to the large confidence intervals given by the variance of data. Nonetheless, the confidence interval in the global difference is shortened because the differences from all questions are combined (whether significant or not), contributing to reducing the variance of data in the global accuracy score analysis and to a highly significant value for the global difference. The more questions are included in the analysis, the more likely it is to detect significant global differences.

Figure 18 provides the difference of accuracy scores between conditions per group of questions. Table 18 provides detailed information on the difference of accuracy scores between conditions for each group of questions. The questions were divided into 4 groups according to their type, as follows: Group 1) questions 1/2/3 of egocentric, depth, fixed estimation; Group 2) questions 4/5/6 of interobject, depth, fixed estimation; Group 3) questions 7/8/9 of interobject, horizontal, exploration estimation; Group 4) questions 10/11/12 of interobject, vertical, exploration estimation. Conditions, group of questions, and the interaction between them were included as predictor variables of accuracy scores (multiple regression model with three predictor variables).

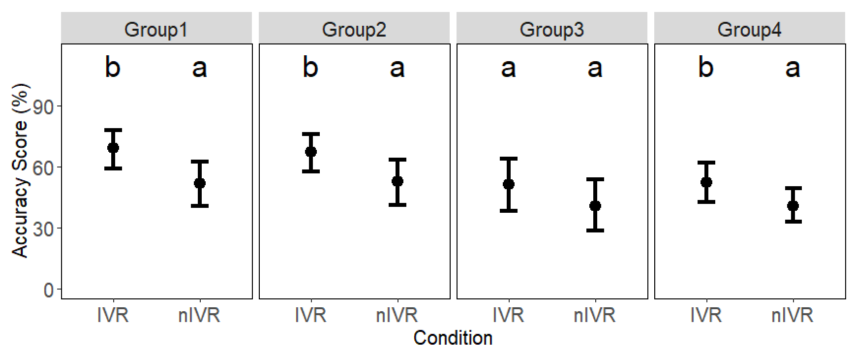


Figure 18. Accuracy scores between conditions, per group of questions

Table 18. Accuracy scores between conditions, per group of questions

Group of questions	Condition	Accuracy Score	Difference	Standard Error	Lower confidence limit	Upper confidence limit	Group	P value
Group 1	niVR	52%	17%	5%	41%	62%	a	**0.002**
	IVR	69%		4%	59%	78%	b	
Group 2	niVR	53%	15%	5%	41%	64%	a	**0.016**
	IVR	68%		4%	58%	76%	b	
Group 3	niVR	41%	10%	6%	29%	54%	a	0.156
	IVR	51%		6%	38%	64%	a	
Group 4	niVR	41%	12%	4%	33%	49%	a	**0.032**
	IVR	53%		4%	43%	62%	b	

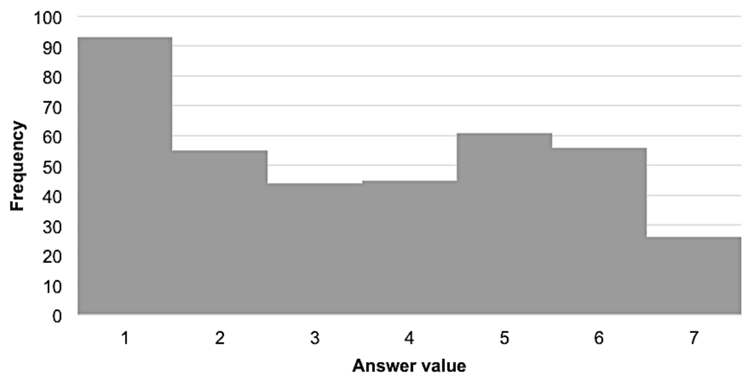
Figure 18 shows that participants did significantly better in IVR than in niVR in egocentric depth estimation at fixed position (Group 1, $p = 0.002$), in interobject depth estimation at fixed position (Group 2, $p = 0.016$), as well as in interobject vertical distance estimation at allowed navigation (Group 4, $p = 0.032$). Table 19 summarizes the results of this section.

Table 19. Differences per question and per group of questions

Category	Dimension	Exploration	Questions	Group of questions	Differences between conditions, per question	Differences between conditions, per group of questions
Egocentric distance estimation	Depth	Not allowed (fixed position)	1, 2, ***3***	1	Q3, in favor of IVR	In favor or IVR
Interobject distance estimation	Depth	Not allowed (fixed position)	4, ***5***, 6	2	Q5, in favor of IVR	In favor or IVR
	Horizontal	Allowed	7, 8, 9	3	-	-
	Vertical	Allowed	10, 11, ***12***	4	Q12, in favor of IVR	In favor or IVR

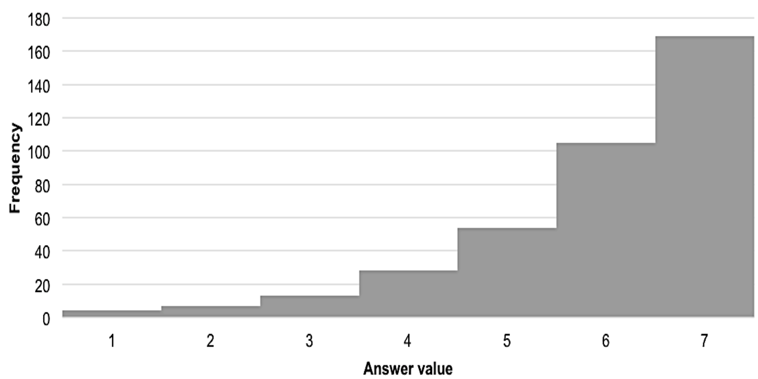
5.2 H2 – Analysis of Presence

The assessment of presence reported in the VR modes involves computing *presence scores*. A participant's presence score is the mean value across ten scores keeping consistency with the instruments of Usoh et al. (2000). Figures 19 and 20 provide histograms and descriptive statistics on the frequency of presence response values (1 to 7) in the niVR and IVR conditions, respectively.



Value	Frequency	Mean
1	93	3.52
2	55	**SD**
3	44	2.01
4	45	**Median**
5	61	3
6	56	**Mode**
7	26	1

Figure 19. Histogram of presence response values in the niVR condition



Value	Frequency	Mean
1	4	5.93
2	7	**SD**
3	13	1.33
4	28	**Median**
5	54	6
6	105	**Mode**
7	169	7

Figure 20. Histogram of presence response values in the IVR condition

5.2.1 Significance Test

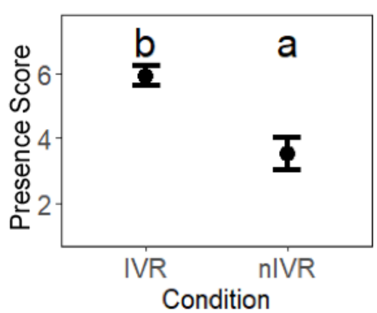
The presence score per participant or per condition is a quantitative continuous variable representing the average of ten numbers that can range from 1 to 7 (a presence score can be any real number between 1 and 7). Therefore, the data set comprising presence scores can be approximated to a normal/Gaussian probability distribution. The total number of presence responses per condition is 380 (38 participants *x* 10 questions).

Similarly to the analysis of 3D perception, three main relationships were tested: a) the difference of global presence scores between conditions, b) the difference of global presence scores across questions, and c) the difference of presence scores between conditions, per question.

The estimated parameters of the adjusted models were extracted along with their respective 95% confidence interval defined by lower and upper confidence boundaries. The letters above each group in the following charts represent the comparisons. Groups that do not share the same letters are significantly different from each other (0.05 of significance). In the analysis per questions, Tukey's correction was applied on the *p* values to avoid inflation of type I error due to multiple comparisons.

5.2.1.1 Difference of global presence scores between conditions

Conditions were included as predictor variables of presence scores and the significance of the difference of global presence scores between conditions was tested (simple linear regression). Figure 21 provides a chart showing the global presence scores in each VR mode, along with a table with detailed information on the scores. The global presence scores differ between IVR and niVR conditions ($p < 0.001$), with higher values found in the IVR condition. Therefore, we refute the null hypothesis of no difference in presence between VR modes.



Condition	niVR	IVR
Presence Score	3.52	5.93
Standard Error	0.22	0.14
Lower confidence limit	3.03	5.61
Upper confidence limit	4.01	6.25
Group	a	b

Figure 21. Global presence scores in the VR modes

5.2.1.2 Difference of global presence scores among questions

Differences of global presence scores among questions were also tested. This helps in the determination of the efficacy and validity of the PQ questionnaire, as it consists of an adapted version based on previous instruments. This analysis encompasses both conditions and focuses on evaluating how each question did in measuring the dependent variable. In this case, PQ questions were included as predictor variables of the presence scores (simple linear regression). Figure 22 shows the global presence scores per question. Table 20 provides detailed information on the global presence scores measured by each question.

Pairwise comparisons (all possible pairs of questions) show that the global presence scores are different within several pairs of questions. The global presence scores of Q6, Q7, and Q9 are the lowest among all questions and also significantly lower than the scores of: Q1 (Q1xQ6, Q1xQ7, Q1xQ9, $p < 0.001$), Q2 (Q2xQ6, Q2xQ7, Q2xQ9, $p < 0.001$), Q3 (Q3xQ6, Q3xQ7, Q3xQ9, $p < 0.001$), Q4 (Q4xQ6, Q4xQ7, Q4xQ9, $p < 0.001$), Q5 (Q5xQ6, Q5xQ7, $p < 0.001$; Q5xQ9, $p = 0.003$), and Q8 (Q8xQ6, Q8xQ7, Q8xQ9, $p < 0.001$). The global presence scores of Q6 and Q7 are also significantly lower than the score of Q10 (Q6xQ10, $p = 0.010$; Q7xQ10, $p = 0.006$). Questions 1, 2, 3, 4, 5, 8 and 10 do not differ among each other. Question 7 is the new question proposed by this study.

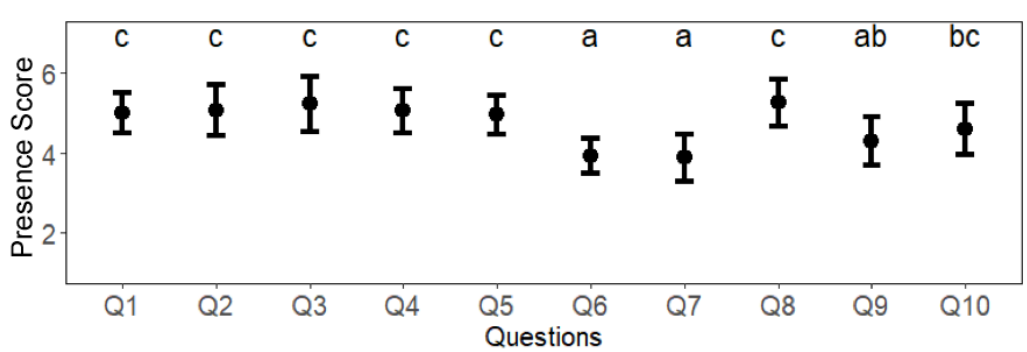


Figure 22. Global presence scores per question

Table 20. Global presence scores per question

Questions	Presence score	Standard Error	Lower confidence limit	Upper confidence limit	Group
Q1	4.99	0.18	4.48	5.49	c
Q2	5.08	0.23	4.44	5.72	c
Q3	5.22	0.25	4.53	5.92	c
Q4	5.05	0.20	4.49	5.62	c
Q5	4.96	0.18	4.47	5.46	c
Q6	3.92	0.16	3.47	4.37	a
Q7	3.87	0.21	3.27	4.47	a
Q8	5.26	0.21	4.67	5.86	c
Q9	4.28	0.22	3.67	4.88	ab
Q10	4.61	0.23	3.96	5.25	bc

The results indicate a slight inconsistency in the level of difficulty across PQ questions. Question 7 – the one proposed by this study – was able to detect a global presence score of the same magnitude to those detected by questions 6 and 9, which were found in the literature and also differ from the remaining questions. Therefore, Q7 performed well in comparison to presence questions found in the literature.

5.2.1.3 Difference of presence scores between conditions, per question

This analysis looks into the difference of presence scores in the interaction between conditions and questions (Figure 23). It aims at examining the extent to which each question contributed to the difference of global presence scores between conditions. Table 21 provides detailed information on the difference of presence scores between conditions for each question. Conditions, questions, and the interaction between them were included as predictor variables of presence scores (multiple regression model with three predictor variables). The presence scores differ between IVR and niVR conditions for all questions, with higher values found in the IVR condition ($p < 0.001$, for all questions).

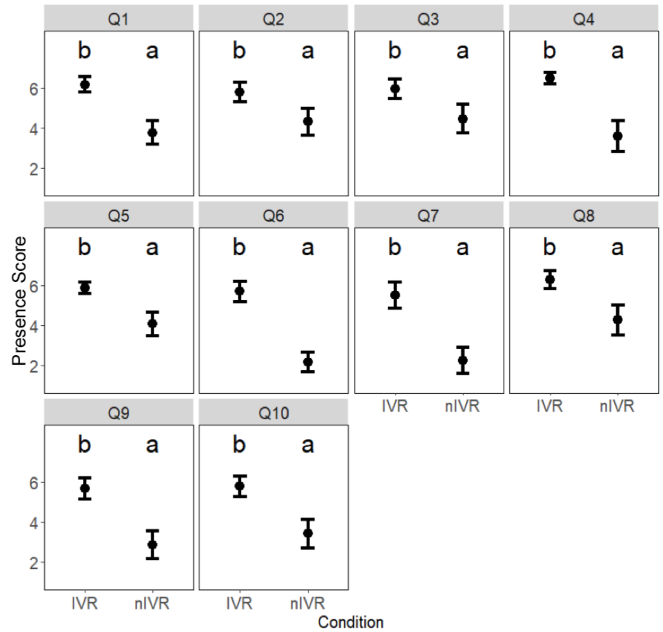


Figure 23. Presence scores between conditions, per question

Table 21. Presence scores between conditions, per question

Questions	Condition	Presence Score	Difference	Standard Error	Lower confidence limit	Upper confidence limit	Group	P value
Q1	niVR	3.79	2.39	0.26	3.21	4.37	a	**< 0.001**
	IVR	6.18		0.17	5.80	6.56	b	
Q2	niVR	4.34	1.47	0.30	3.67	5.01	a	**< 0.001**
	IVR	5.82		0.22	5.33	6.30	b	
Q3	niVR	4.47	1.50	0.32	3.75	5.19	a	**< 0.001**
	IVR	5.97		0.22	5.47	6.48	b	
Q4	niVR	3.61	2.89	0.35	2.82	4.39	a	**< 0.001**
	IVR	6.50		0.12	6.23	6.77	b	
Q5	niVR	4.05	1.82	0.27	3.46	4.65	a	**< 0.001**
	IVR	5.87		0.13	5.58	6.16	b	
Q6	niVR	2.16	3.53	0.22	1.68	2.64	a	**< 0.001**
	IVR	5.68		0.22	5.19	6.18	b	
Q7	niVR	2.24	3.26	0.29	1.59	2.88	a	**< 0.001**
	IVR	5.50		0.29	4.86	6.14	b	
Q8	niVR	4.26	2.00	0.33	3.52	5.01	a	**< 0.001**
	IVR	6.26		0.20	5.82	6.71	b	
Q9	niVR	2.87	2.82	0.30	2.19	3.55	a	**< 0.001**
	IVR	5.68		0.23	5.16	6.20	b	
Q10	niVR	3.42	2.37	0.32	2.70	4.14	a	**< 0.001**
	IVR	5.79		0.23	5.28	6.30	b	

As revealed in the previous analysis, Q6 and Q7 yielded significantly lower global presence scores (both conditions considered) in relation to the other questions, meaning that participants reported significantly lower presence with respect to the questions' statements. Nonetheless, this analysis revealed that, at the same time, Q6 and Q7 also detected the largest significant differences between conditions in favor of IVR (3.53 and 3.26 respectively). In other words, although participants reported lower presence levels with those questions, those were also the questions that detected the largest differences in presence levels between conditions. It is worth noting that Q7 is the new question proposed by the researcher. The aforementioned results reinforce the good performance of this question.

5.3 H3 – Analysis of association between 3D Perception and Presence per Condition

This analysis examines whether dependent variables are associated within the VR conditions. In statistical terms, two variables are associated if there is a significant relationship between them. It should be noted that a significant association does not necessarily mean that changes in one variable cause changes in the other. In some cases there is a hidden variable, also called intervening variable (a type of confounding variable) that acts as the underlying cause of those changes (Lazar et al., 2017). The expected association tested is unidirectional, that is, 3D perception affecting presence.

For this analysis, while the presence score continues to be treated as a quantitative continuous variable, the accuracy score is treated as a percentage, i.e., as a quantitative continuous variable as well, instead of as a categorical nominal variable as in the previous analyses. Therefore, both variables assume normal/Gaussian probability distributions. Conditions and accuracy scores were included as predictor variables of presence scores (multiple regression model with two predictor variables). Figure 24 shows the linear relationship between accuracy scores and presence scores in each condition. There is no significant association between accuracy scores and presence scores in the conditions (IVR slope: r = 0.00495, p = 1.0; niVR slope: r = - 0.175, p = 0.3). Therefore, we fail to refute the null hypothesis of no relationship between 3D perception and presence in the VR modes.

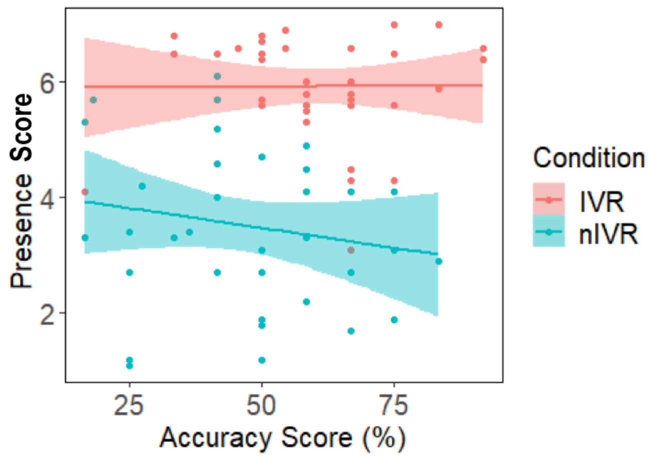


Figure 24. Relationship between 3D perception and presence in the VR modes

In Figure 24 it is possible to see the data points, which correspond to each of the 38 participants. It becomes clear from the figure that there is no pattern in any condition. For instance, while in the IVR condition most participants reported great levels of presence, some of them achieved poor 3D perception while others achieved the opposite.

5.4 Analysis of Individual Factors & Order Effects on 3D Perception and Presence

This analysis verifies the influence of individual factors (confounding variables) and sequence/order of presentation of conditions on the 3D perception and presence variables.

Confounding variables and order of presentation were included in the models to check if they would distort the effects of conditions and questions on the dependent variables. Two separate multiple regression models were adjusted for each response variable, i.e., one model predicting presence and a separate one for 3D perception. In summary, each model included conditions, questions, and sequence (controlled variables) along with confounders, predicting a single dependent variable. It should be noted that confounding variables are included as predictor/independent variables in the models. However, because these were actually not controlled/manipulated, any effects eventually found are restricted to this study's sample, not being generalizable to the study's population (only the effects of manipulated variables are generalizable) (Pujoni, 2019).

Effects of Individual Factors

It is only possible to include confounding variables in the model that are independent from each other. Therefore, the first step of this analysis was to identify which confounders are significantly associated with each other so that a single factor from a pair or cluster of associated factors could be selected and included in the model as an independent variable. Reasonable associations such as age and educational level, or age and experience in design review do not compromise interpretations. However, the effects of factors that are associated with each other should not be interpreted separately. For instance, if educational level is associated with age, and it affects a dependent variable, it would not be possible to attribute that effect to educational level only, as it could also have been due to age.

The individual factors analyzed are nine: age, gender, educational level, bachelor's major, current major, experience in design review, experience with 3D virtual environments, familiarity with the experiment environment, and spatial ability. The analysis of association revealed ten significant associations among eight factors (gender – the ninth factor – is not associated with any other factor). Four factors (besides gender) were then selected and included in the models. The decision on which factors to select within a pair or cluster of associated factors is made based on their relevance in light of the general research goals. The selected confounders were deemed as better predictors of the response variables for different reasons. For instance, bachelor's major, which is interpreted as one's main professional role or original field of practice, is selected over educational level because the study's population of interest consists of professionals whose roles are primarily defined by their bachelor's majors (architect, civil engineer, etc.), not their educational levels. Another example is the decision between experience with 3D virtual environments and spatial ability. While both could affect 3D perception in virtual environments, the former is a rather subjective, self-reported response, whereas spatial ability was measured through a more objective method and hence is seeing as more reliable.

The inclusion of individual factors and order of presentation in the models revealed the significance of their effects on the dependent variables. Significance values are provided in Table 22 below.

Table 22. Effects of individual factors and sequence on the dependent variables

	Dependent variable			
	3D perception		Presence	
Confounding variables	X^2	P value	X^2	P value
Bachelor's major	89098.1	**< 0.001**	0.0	1.0
Experience in design review	3.8	0.052	1.6	0.204
Familiarity with the experiment environment	13.2	**0.004**	20.1	**< 0.001**
Spatial ability	0.0	0.959	2.5	0.116
Gender	0.0	0.935	9.5	**0.002**
Controlled variables				
Conditions	20.0	**< 0.001**	153.5	**< 0.001**
Questions	50.4	**< 0.001**	177.5	**< 0.001**
Sequences	0.0	0.954	7.6	**0.006**

The significant individual factors for 3D perception were bachelor's major ($p < 0.001$) and familiarity with the experiment environment ($p = 0.004$). For presence, significant confounders were familiarity with the experiment environment ($p < 0.001$) and gender ($p = 0.002$). The significant effects of confounding variables did not alter the significance of the effects of conditions and questions on the dependent variables. That is, even under significant influence of individual factors, the difference between conditions remains significant. In other words, conditions remained a significant factor in the explanation of 3D perception and presence. The effects of conditions on the dependent variables were so strong that these remained significant even in the presence of (or "controlled by") those confounding variables.

Order Effects

In order to check if sequence of presentation of conditions had any effect on the dependent variables, order/sequence was also included as a predictor/independent variable in the models. Order effects are any significant differences in the dependent variables between groups that received treatments in different order (group 1 – participants exposed to sequence 1 *vs.* group 2 – participants exposed to sequence 2). The order of presentation of conditions was manipulated as described in item 4.2.7. When included in the models as an independent variable along with conditions, questions, and confounders, the effects of conditions and questions on the dependent variables remain significant (Table 22). The significant association between sequence and presence (p = 0.006) did not alter the significance of the effects of conditions and questions on the presence response. In other words, conditions remained a significant factor in the explanation of 3D perception and presence. In sum, no order effect on 3D perception was detected, and the order effect on presence did not alter the effect of conditions on this variable.

Additional analyses were conducted to check if there are significant differences between the responses of the two groups of sequence, that is, if differences in 3D perception and presence between conditions are different between groups 1 (niVR-IVR) and 2 (IVR-niVR). The first analysis looks into the difference of accuracy scores in the interaction between conditions and sequences (Figure 25). Table 23 provides detailed information on the difference of accuracy scores between conditions for each group of sequence. Conditions, sequences, and the interaction between them were included as predictor variables of accuracy scores (multiple regression model with three predictor variables). No difference between conditions was detected in the group 2 whereas there is a difference between conditions in the group 1 (21%), meaning that the differences within each group are different between groups (p = 0.0062).

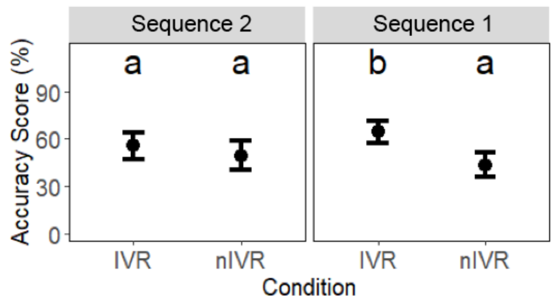


Figure 25. Accuracy scores between conditions, per group of sequence of presentation

Table 23. Accuracy scores between conditions, per group of sequence of presentation

Group of Sequence	Condition	Accuracy Score	Difference	Standard Error	Lower confidence limit	Upper confidence limit	Group
1	1st niVR	43%	21%	3%	36%	51%	a
	2nd IVR	65%		3%	58%	71%	b
2	2nd niVR	50%	6%	4%	40%	59%	a
	1st IVR	56%		4%	47%	64%	a

Likewise, the second analysis looks into the difference of presence scores in the interaction between conditions and sequences (Figure 26). Table 24 provides detailed information on the difference of presence scores between conditions for each group of sequence. Conditions, sequences, and the interaction between them were included as predictor variables of presence scores (multiple regression model with three predictor variables). Differences between conditions were detected in both groups 1 and 2; however, the difference of group 2 (2.97) is larger than the difference of group 1 (1.84). Therefore, the differences within each group are different between groups (p = 0.0010).

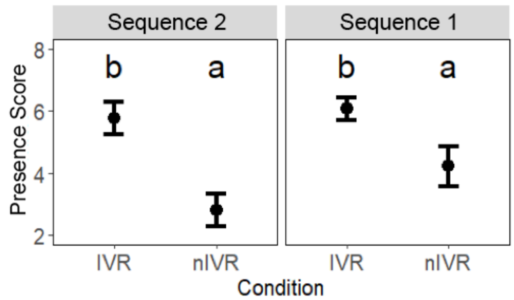


Figure 26. Presence scores between conditions, per group of sequence of presentation

Table 24. Presence scores between conditions, per group of sequence of presentation

Group of Sequence	Condition	Presence Score	Difference	Standard Error	Lower confidence limit	Upper confidence limit	Group
1	1st niVR	4.24	1.84	0.29	3.59	4.88	a
	2nd IVR	6.08		0.16	5.72	6.44	b
2	2nd niVR	2.81	2.97	0.24	2.28	3.33	a
	1st IVR	5.77		0.23	5.26	6.29	b

CHAPTER 6

DISCUSSIONS

The demographics data suggest that the nonprobability-based sample is a good representative of the study's population of interest (current and future workforce). Considering the sampling process, context, and constraints, expected significant associations among individual factors were confirmed such as between age and educational level, age and experience in design review, and current major and familiarity with the experiment environment.

H1 – Analysis of 3D Perception

Hypothesis 1 is associated with the main question in this dissertation. This question compared the ability of two distinct VR systems in conveying the three-dimensionality of a BIM-based architectural model.

The comparison of accuracy scores between VR modes indicates that IVR technology can better reproduce a user's 3D perception in the real world than non-immersive VR. Participants achieved a significantly higher accuracy score in the IVR condition, meaning that it provided the most similar 3D perception to that obtained in the real world. In order words, IVR can be deemed the technology that offers the most realistic 3D perception experience of built spaces. When using the immersive technology participants had a better 3D perception of the architectural representation in comparison to their perception using the conventional workstation. The IVR system allowed users to perceive three-dimensional features more accurately.

IVR appears to be the most appropriate system for tasks that benefit from accurate representation and communication of three-dimensional spaces. The perception of three-dimensionality was operationalized in this study as one's distance estimates. The greater accuracy in distance estimation using the immersive platform implies a better understanding of the three-dimensional configuration of the space depicted. Whether a user is able to better estimate distances – when one's virtual estimates are closer to real-world estimates – it means that she/he better understood the three-dimensional configuration of the virtual model. An enhanced 3D perception can be interpreted as a better understanding of the three-dimensional relationships of the architectural representation, meaning that within the immersive environment geometric information "makes more sense" and can be better assimilated by the observer. In summary, the representation and communication of three-dimensional information are leveraged with support of the immersive system. Ultimately, a better understanding of the virtual model is expected to benefit design review.

Results also provide insights regarding distance underestimation in virtual environments. Previous studies found that distances appear more compressed in virtual environments (15% of compression) than they do in the real world (8% of compression) with respect to actual dimensions (Gooch and Willemsen, 2002; Thompson et al., 2004; Renner et al., 2013). This study did not verify whether judgments in the virtual environments were over or underestimated in relation to real-world estimates. It only examined how often judgments in the virtual and real worlds would match (within predefined ranges of distance values). The IVR estimates were found to match real-world estimates more often than niVR estimates did (hit occurrence). This result alone means that distance distortion in IVR is less likely, which leads to the expectation that underestimation might occur to a lesser extent than it occurs in niVR. It should be noted that a realistic virtual environment is one that reproduces the 8% of underestimation that happen in real life. Thus, distance compression in IVR is possibly closer to those 8% of compression in the real world, which makes it the most realistic simulation in terms of distance perception. However, the experiment was not designed to quantify the difference of underestimation between virtual and real environments (15% – 8% = 7%). Regardless, findings still suggest that it is probably smaller between IVR and real world.

It should be noted that it makes no difference in this study if and the extent to which people underestimate distances in the real world. The experiment was designed to take underestimation in the real world into account regardless of its magnitude. The estimates in the VR modes were "standardized" over real-world estimates and only after this procedure the VR modes could be compared between each other.

Participants did significantly better in IVR than in niVR in three of four types of distance judgments – egocentric and interobject depth estimation at fixed position, as well as in interobject vertical distance estimation at allowed navigation. This result suggests that, in general, people estimate distances more accurately in the immersive environment regardless of the vantage point circumstances (allowed exploration or fixed position) and distances being estimated (egocentric or interobject).

As indicated by the analyses of difference of accuracy scores among questions, and between conditions per question, familiar architectural elements such as doors and staircase steps may also help in 3D perception, regardless of the VR mode. Participants strongly relied on their knowledge about such elements in the environment, which ended acting as major depth cues in both conditions.

Although the technological and representational factors in charge of promoting better 3D perception in the immersive system were not investigated separately in this study, stereopsis – a visual cue specific to the immersive condition – is most likely among them. Other system properties such as field of view and interaction devices may also have affected 3D perception. However, it is impossible to isolate stereopsis from those due to the very nature of VR systems utilized in this study. While Kalisperis et al. (2006) and Zikic (2007) focused on measuring the effects of technological characteristics of immersive systems such as display features (e.g., stereoscopy, screen size, field of view) and representational aspects (e.g., level of realism, level of detail) on perception and presence, this study provides the global efficiency of VR systems given all their fixed factors. Regardless, results suggest that the immersive environment provided critical depth cues to enable a more realistic depth perception, as opposed to depth cues available in the niVR condition. Since featural cues (representational aspects) were kept the same across conditions, stereopsis is certainly one of the cues responsible for that difference. While occlusion and perspective are important depth cues (Cutting and Vishton, 1995), results suggest that stereopsis may also play a crucial role in depth judgments in virtual environments, confirming the expectations of England et al. (1992). In a scenario where stereopsis is provided in addition to occlusion and perspective (as it was in the immersive condition) participants report more accurate depth perception.

H2 – Analysis of Presence

This question compared the ability of two distinct VR systems in generating sense of presence. Results suggest that IVR technology offers greater levels of presence, meaning that it provided more immersive experiences and confirming the findings of Witmer and Singer (1998), Kalisperis et al. (2006), Zikic (2007), and Castronovo et al. (2017). IVR appears to be the most appropriate system when a VR-supported task benefits from immersion and involvement, such as design review. Greater levels of presence are expected to enhance a user's ability to perform visual search and understand the displayed information while interacting with a design representation hence facilitating the identification of design issues and ultimately benefiting the design review process.

Significant differences in some questions were expected, but not for all questions as revealed by this analysis. In every question, presence in IVR was significantly greater than in niVR. The analysis shows that large-sized significant differences in presence scores between conditions could have been detected with a single question (any question), which reinforces the fact that if power analysis had been done based on the presence score variable, the sample size would have been much smaller, thus compromising the power to detect differences in the other response variable of 3D perception.

H3 – Analysis of association between 3D Perception and Presence per Condition

Identifying and characterizing the factors that affect presence is critical for the development of increasingly effective VR systems. Presence might be influenced not only by the platform characteristics (technological factors) but also by user characteristics (human factors). In the quest for the identification of these factors, the researcher hypothesized that there could be a relationship between 3D perception and presence in the virtual environments. The expected association tested is unidirectional, that is, 3D perception affecting presence.

On one hand, results indicate that 3D perception may not be among the factors affecting presence, as opposed to the expectations of Steuer (1992), Bertol (1997), and Witmer and Singer (1998). On the other hand, the inexistence of a significant association endorses the argument of Interrante et al. (2008) who stated that people might still perform well in estimating distances in virtual environments that do not offer the conditions for great levels of presence. The results also coincide with previous studies suggesting that accurate distance estimation may not be necessarily an evidence of great levels of presence (Thompson et al., 2004; Kalisperis et al., 2006; Interrante et al., 2006; Renner et al., 2013).

In summary, presence may not have any relationship with the three-dimensional realism of a scene or experience. As suggested in the literature, people may still feel strongly present in non-realistic environments whether these are places depicted in two-dimensional films, the mental depiction of a place while reading a novel, or even places visited in dreams.

Analysis of Individual Factors & Order Effects on 3D Perception and Presence

In order to check for the effectiveness of measures taken to control individual factors with the experimental design, the effects of these confounding variables on 3D perception and presence was verified in the data analysis. The effects of conditions and questions on the dependent variables remained significant even in the presence of (or "controlled by") confounding variables. In other words, conditions remained a significant factor in the explanation of 3D perception and presence. Also, order effects were either not detected or did not impact the effects of conditions and questions, confirming that the randomization and counter balancing procedures were effective.

In previous studies that focused on the influence of individual factors on presence (e.g., Stanney et al., 1998; Nowak et al., 2008) such factors were manipulated so that results were generalizable to the population. The associations detected in this research, however, are restricted to its sample, not being generalizable to the study's population. Gender, familiarity with the experiment environment, and sequence of presentation were found to significantly affect presence. Participants exposed to sequence 1 (niVR, IVR) reported different presence levels than participants exposed to sequence 2 (IVR, niVR), both conditions combined. Moreover, the differences between conditions are different between groups of sequence. The significant effects of sequence on presence raise the need for a closer examination because sequence and gender are significantly associated, i.e., gender is not balanced between groups of sequence. It is impossible to attribute the effect of sequence on presence to sequence only, as it could have been due to gender. For instance, while sequence 2 shows a significantly larger difference between conditions in comparison to group 1, it also has significantly more female participants.

These findings could also suggest the existence of an unappreciated variable possibly associated with the feeling of amusement when exploring the immersive environment experienced by participants exposed to the non-immersive simulation first (sequence 1), which was possibly stronger than the opposite situation and may have influenced their responses in the second condition. Kuliga et al. (2015) also noticed a possible relationship between performance in VR environments and presentation order. Thus, future studies may consider controlling for likely effects of gender and order of presentation through manipulation of these factors in the experimental design.

In a previous study conducted by the researcher (Paes et al., 2017), familiarity with the experiment environment was found to have no influence on spatial perception. However, in this research it did impact 3D perception (and presence). 3D perception was influenced by familiarity with the experiment environment and bachelor's major. The interpretation of the effect of bachelor's major on 3D perception is problematic due to its several categories with a single participant. It is possible that an intervening variable is affecting this association (perhaps current academic major, which is associated with familiarity).

While Kovac (1989) suggests that spatial ability is positively correlated with users' performance in spatial tasks, both spatial ability and experience in design review did not show any significant effects on the dependent variables, contrary to the researcher' expectations. However, this finding is consistent with the expectations of Witmer and Singer (1998) and Calderon-Hernandez et al. (2019).

Practical Significance

Cognitive processes are in charge of human performance; therefore it is important that VR effectiveness assessments take into consideration the cognitive processes that underlie performance in virtual environments. In conformity with the main references in the fields of HCI and cognitive psychology, effectiveness analyses of virtual environments should be conducted on measures of user performance.

IVR effectiveness over traditional VR in design review has been repeatedly reported in the past. However, the vast majority of such studies demonstrating IVR implications on productivity and task performance often delivered a rather anecdotal report of users' experiences with IVR systems, disregarding the eventual cognitive improvements provided by such technology. A fundamental step for IVR to be considered an effective technology for collaborative design review is to understand whether and to what extent it enhances users' cognitive abilities, that is, their performance in obtaining and assimilating the spatial relationships depicted.

In this context, this study expands the characterization of the benefits repeatedly reported in the past. To "assess the *extent to which* IVR offers better support to design review compared to non-immersive VR platforms" means to examine the phenomenon that is causing people to better understand a 3D model using IVR technology. This research provides a deeper analysis of what is actually happening when a user reports a better understanding of a project's 3D model through immersive visualization. As per Wann and Mon-Williams (1996), the effectiveness of virtual environments is dictated by perceptual criteria. Thus, this study focused on the characterization of the perceptual experiences that underlie the increased IVR effectiveness over traditional VR.

The factors that lead users to report increased effectiveness are not directly related to the VR system per se, but to the cognitive processes that have been enhanced by the technology (Figure 2). Evidence show that when people report that IVR is more efficient in the design review, this is related to a better understanding of the spatial relationships of a 3D model. Based on the literature, this study hypothesized that reported benefits are associated with enhanced 3D perception of the design representation and greater involvement in the review activity. Hypotheses tests were conducted and indicated that the IVR system does provide enhanced 3D perception and greater levels of presence. This study did not cover technological factors that may contribute to those benefits. However, it provides evidence that 3D perception and presence in the review of 3D models are improved using IVR technology compared to non-immersive VR systems.

In this context, practical significance of the differences between technologies lies on the relevance of such differences in light of the design practice, and it is given by the high probability to find large improvements in the study's population. Although small-sized effects may exist, these were deemed not relevant to the ultimate purpose of this study – to verify the existence of effects of immersive visualization that would significantly benefit design review. Thus, *a priori* power analysis was conducted based on a meaningful and considerably high difference in 3D perception between technologies, which ensured an adequate sample size for a high probability (80% power) to detect existing differences of 33% and up. One can expect to find such improvements in the population as well. In sum, results provide practical significance to the extent that these indicate a high probability to find large-sized improvements in the population, which is relevant to the practice. The ability of IVR technology in providing users with significantly better 3D perception is expected to improve the understanding of 3D models and, consequently, collaborative design review.

6.1 Challenges Encountered & Lessons Learned

1. Studies of this nature should take into account the time, sampling, and resource limitations when designing and conducting experiments involving human subjects. When no financial or other type of compensation is offered, researchers rely on the availability and willingness of potential participants and the recruitment process may take longer than expected. Also, one must pay attention to undesirable effects of fatigue, which are particularly detrimental to within-subject studies. Experiments that involve considerable cognitive and/or physical effort should not take longer than a reasonable time per participant. This is subject to the tasks being tested, the type of data and the collection method, and even the physical conditions of the

experiment location. How participants would feel over the experiment session should be verified beforehand through pilot experiment trials. Otherwise, fatigue effects may jeopardize participants' performance during actual data collection.

2. When dependent variables involve significant cognitive functions, individual differences are expected to largely affect the outcomes. Such individual differences are better controlled in a within-subject design, which excludes the variance between subjects due to those differences in the comparison of effects of different conditions, since each participant serves as her/his own equivalent in the comparison. Nonetheless, within-subject experiments must be carefully designed to ensure that the benefits of such a design (smaller sample size, greater power, etc.) outweigh potential drawbacks (practice/learning/carry-over, fatigue, and expectancy effects). These order effects can be controlled through random and balanced assignment of participants to experimental conditions, as well as by sorting the order of questions of data collection instruments used in consecutive conditions.

3. Because perception and thought are produced at the unconscious level, validity studies must be conducted using the most objective assessment methods possible, as well as techniques to counterbalance the inherent subjectivity of self-reported judgments about space. Standardizing virtual-world estimates over physical-world responses can minimize such subjectivity (Higuera-Trujillo et al., 2017). When the purpose of a virtual environment is to reproduce human cognition and behavior in the real world, the effectiveness of such technology is determined through a comparison of user performance in the virtual setup against performance in the real world (Gooch and Willemsen, 2002). Thus, in studies that compare distance estimation performance between different virtual environments, estimates in the physical world should be used to standardize virtual estimates (Paes et al., 2017). Naturally, such studies would automatically assume that the cognitive mechanisms that govern the processes of distance perception are the same in all

conditions, such that comparing distance judgments in virtual environments to those made in physical environments can be deemed a valid procedure.

4. *A priori* power analysis is also indispensable to determine the adequate sample size to provide a test with sufficient power to detect existing differences. The effect size of interest – the magnitude of the difference between treatments – should be carefully determined during power analysis. An estimate of the difference expected could be done through research literature, pilot study, expert judgment, or educated guessing. Defining a meaningful effect size is important for establishing the practical significance of effects observed. However, it should be noted that the meaningfulness of an effect size is subjective and varies largely across studies. Although Cohen (1988) provides standard values of effect magnitudes for studies in the social and behavioral sciences, effect sizes should be determined on a case-by-case basis, as it is subject to the response variables, aims, instruments, experimental design, and even sampling factors of a study.

5. In a preliminary study conducted by the researcher (Paes et al., 2017), many performance comparisons were based on individual characteristics (age, gender, experience, etc.). However, since those individual characteristics were not manipulated/independent variables, the analyses of perception performance (dependent variable) per individual factor were not appropriate. If individual factors had been treated as independent variables, the researchers would have needed to recruit a sufficiently large number of participants within each sample subgroup defined by those individual factors, that is, at least 30 male participants, 30 females, 30 novices, 30 experts, and so on. In this dissertation, individual factors were not treated as independent variables but were controlled by testing their impacts on 3D perception and presence to check for any effects on the response variables.

6. Regardless of the strategies adopted to mitigate the problem of systematic similarities between conditions, participants are not likely to learn over successive exposure to experimental conditions because these are not exactly similar. In this study, the extent to which early trials induce a response bias that influence performance on later trials is reduced by asking participants to "rethink their answers" since stimulus varies across conditions (Thompson & Campbell, 2004). Participants are not told the correct answer after being exposed to a condition, that is, they do not know if their guesses in early trials are right or wrong so that there is no useful information that could "prime" or compromise their decisions in the next condition. In other words, participants could not use their guesses in the previous condition as reference to try to guess more accurately in the next condition, and they are made aware of that. In this study, being successively exposed to stimulus does not make a participant more likely to be more precise in the next condition simply because stimuli are fundamentally not similar (different media cause different stimulus). Participants do have the chance and are actually oriented to reevaluate their responses, and this is precisely the goal of the experiment. Even though they may recall their answers in early trials, this does not make any difference because they are aware that stimulus varies across conditions so that their previous responses should not be assumed as correct.

7. Initially, the strategy suggested by Lazar et al. (2017) of transforming 3D perception data into percentage values so that they are normally distributed and adequate for ANOVA or t tests was considered. In this study there are 12 measures of 3D perception (hit/error values) per participant (the outcomes of each of the 12 questions), in a total of 456 observations per condition (12 measures *x* 38 participants). Naturally, the process of transforming these data into individual percentage values per participant would involve collapsing those 12 observations into a single mean score per participant. ANOVA and t tests would average the resulting 38 mean scores into global means per condition (global accuracy score), and then calculate the significance of the

different between these global means. This process would represent a large loss of data, reducing 456 observations into 38 observations, i.e., decreasing the sample by 12 times.

This process could have been performed for the analysis of the significance of the difference of global accuracy scores between conditions in item 5.1.1.1. However, it was not performed because t test or ANOVA-based analysis of a single normally distributed global accuracy score per condition would compare the mean across those 38 values between conditions, which, again, would represent a huge data loss. This strategy would disregard the 456 observations in a condition (12 *x* 38), computing the mean value across 38 mean scores in each condition, and then comparing the global means between conditions regardless of the data set from which those means were generated.

The point is that a global mean from a set of 38 scores is not as robust as a global mean from a set of 456 observations. The latter situation is more reliable. Assuming a normal distribution for the data set made of categorical values would disregard that distinction. Collapsing 456 data points into means per participant (transforming categorical into continuous data) and then comparing the global means per condition, as opposed to directly comparing the 456 individual observations per condition, is a strategy that leaves out a great amount of data, decreasing the power of detecting existing differences. In turn, assuming binomial distribution for the 3D perception variable and adjusting regression models for the analysis allows for the examination of the significance of the difference of global accuracy scores between conditions utilizing all 456 observations, increasing the probability of making correct inferences. Assuming binomial distribution, a regression analysis does not compare 38 accuracy scores per participant between conditions, but 456 values between conditions.

In addition, collapsing the observations from 12 questions into a single accuracy score per participant makes it impossible to compute the difference of accuracy scores among questions, since the effect of a question is lost when the set of 12 is transformed into a single mean value. Therefore, the binomial distribution allows for the examination of the differences of accuracy scores among questions as well, through regression analysis.

8. Assuming the probability distribution that better fits a particular data set increases the probability of making correct inferences. In this study, a binomial distribution was considered more adequate for the analysis of the 3D perception categorical variable. The analysis of this binomially distributed variable required the adjustment of several regression models (GEE). Future similar studies should pay special attention to the choice of the most appropriate probability distribution and related statistical methods. In general, it is not recommended to assume normal distribution for categorical variables.

6.2 Contributions

This research contributes to the debate on the effectiveness of IVR technology in the design process providing evidence about the existence and extent of improvements to 3D perception and presence levels of a specific user population, eliciting the effectiveness of immersive representations, and providing insights to the development of more effective IVR systems. Another contribution of this work lies in its methodological innovation. The combination of four methodological characteristics (context, comparative, quantitative, and user-centered approach) allowed for the development of a method able to provide answers to questions little explored in the design domain.

Major innovations over past research include: a) the consideration of a participant's 3D perception in the real world as a reference for establishing the accuracy of a virtual environment (standardization over real-world responses), b) development of an adapted version of a 3D Perception Questionnaire, c) development of an adapted version of a Presence Questionnaire, and d) investigation of the relationship between presence and 3D perception in virtual environments.

The innovative aspect regarding the latter item is that while presence and distance estimation in virtual environments have been investigated separately in the past, the relationship between them has never been examined before although scholars suggest that it exists.

From the initial attempts of Witmer and Singer (1998) to nowadays, avoiding or reducing subjectivity of questionnaires is one of the major challenges in studies of perception and presence in virtual environments. As an improvement from previous comparative methods, and based on studies from the cognitive sciences, a participant's 3D perception in the real world was collected and directly compared to 3D perception in the virtual environments. The experiment was designed to investigate whether a participant's 3D perception in the VR modes deviates from perception in the real environment, regardless of what estimates were provided in the real world. Thus, the method is at the same time: a) not concerned whether real-world estimates are accurate or not in relation to the actual dimensions, and b) taking into account these estimates to standardize VR estimates. In short, the assessment of 3D perception in the VR modes is done by comparing it to 3D perception in the real world. This comparison provides a VR system's effectiveness in terms of its ability to simulate/represent a given space, from the user standpoint.

It should be noted that although this study checks for whether 3D perception in virtual environments would reproduce the performances observed in the real world (whatever this performance was, accurate or not), this is an intermediate step in the experiment. The ultimate goal is to compare the results of this step, regardless of the likely underestimations in all conditions, as indicated in the literature (Gooch and Willemsen, 2002; Thompson et al., 2004; Renner et al., 2013).

Lastly, although developing new measurement instruments was not among the main goals of this research, it may still provide some contributions to this matter since the assessment of the PQ and 3DPQ scales utilized in this study indicated some internal consistency and validity of the instruments. This study utilized sufficiently consolidated and well-accepted instruments in the field of VR research, and examined their effectiveness after adjustments.

CHAPTER 7

CONCLUSIONS

The advent of the perspective technique represented a major game changer in the representation of architecture. Perspective constructions were able to deliver the most accurate representations of three-dimensional architecture artifacts, representing a revolution in the long-running quest for realism of architectural representations. The expression of depth through perspective provided more realistic depictions of reality hence improving the representation of envisioned environments. Recently, the advent of advanced visualization tools such as immersive virtual reality may represent the next revolution in the architectural representation paradigm, leveraging the advantages of stereoscopic, life-size, and interactive visualization to expand the representation of three-dimensionality beyond the boundaries of flat media.

An interdisciplinary research field that integrates cognitive parameters with design technology to improve the interface between humans and computers in design applications may be emerging. In this context, this research provides evidence that immersive virtual reality actually improves the perception of three-dimensionality from virtual architectural models. An IVR-based model provides the most realistic representation of the three-dimensional relationships of an architectural object. Thus, it is expected to facilitate collaborative design review as well, to the extent that professionals involved in the task are more likely to understand the information under evaluation as well as to get involved with the review process. It is the representation format that works best from the user standpoint. However, while perception has been acknowledged as the “missing link” between pictorial realism of VR simulations and presence, this research could not detect such a relationship (between perception and presence) in neither of the virtual systems studied.

It should be noted that a better understanding of the three-dimensionality of a virtual model and greater levels of presence in the simulation do not necessarily imply a better understanding of a designer's intentions nor smarter design solutions. Nonetheless, it would represent a critical step towards it since architectural design is mainly concerned with solving spatial problems. It is reasonable to expect that once the three-dimensionality of a virtual model is better understood and greater levels of presence are reached, more suitable solutions to those spatial problems are likely to arise. However, the cause-effect relationship among enhanced 3D perception, levels of presence, and quality of design solutions could not be demonstrated through this study and may be addressed in future research.

IVR systems have the potential to revolutionize the way industry professionals perform many of their tasks and, in fact, they have already started to impact design environments. User-centered research on IVR systems is relevant to both industry and academia to the extent that it could ultimately lead to the development of novel collaborative IVR environments and methods, tailored to specific users and application contexts. Research in this field will also provide reliable evidence on IVR systems' benefits, effectiveness, and limitations in different contexts of application.

Virtual reality can serve many purposes and, as discussed by Berg and Vance (2016b) and Paes and Irizarry (2019), the level of realism of a simulation is strictly a function of the questions being explored and goals with that simulation. A researcher wanting to investigate interceptive timing behavior might wish to violate Newtonian mechanics so that objects can move in unexpected trajectories (Brookes et al., 2019). In that case, an unrealistic simulation (with respect to the rules of Newtonian mechanics) is necessary. Nonetheless, the power to resemble the world as people see and interact with is precisely what makes IVR technology a promising tool to simulate, predict, and investigate critical implications of architectural design solutions according to the reality that people are able and used to experience.

7.1 Limitations & Future studies

It should be noted that while different equipment may prompt different responses from the ones observed in this study, the focus of user-centered studies in the field should not be on the systems, devices or equipment, but on the user experience with the technology. Because technology is rapidly evolving and equipment becomes easily obsolete, if a validation study of different platforms is applicable, this should be done in light of user experiences. Nonetheless, the particular combination of hardware and software utilized in this study restricts the research findings to its particular technological setup. Also, it is possible that the results obtained in this research are restricted to the particularities of the modeled environment (the Caddell Building lobby), so that when altering spatial properties or analyzing a different space results could change. As future work, it would be interesting to replicate this study using different models.

The level of realism of a simulation is application-dependent. Virtual environments may not need to include certain visual cues if an application context does not require such information. Ultimately, pictorial and physical resemblance to the real world may not be necessary in some contexts. Therefore, future research may look into the visual and interactive features required in immersive simulations for optimal cognitive performance and productivity, that is, explore and identify what visual cues (shading, texture, occlusion, stereopsis, parallax, motion perspective, etc.) and equipment properties (display resolution, latency, field-of-view, interfaces, etc.) would deliver optimal responses (not necessarily presence and perception) and benefit a given application the most (e.g., architectural design, safety training, etc.). One may find out an equation that allows the trade off among visual cues, technological and human factors towards enhancing user performance in a virtual environment. The path towards such an equation starts by understanding the depictive information needs in each application context, which is another entirely independent topic for future research as well.

Regardless of the contributions of this work, research in the AECO-FM domain has yet to demonstrate the benefits of IVR systems to other aspects of the design process as well as to other construction tasks. In this study, the approach was to test whether and to what extent IVR systems were able to deliver more effective representations than traditional systems in terms of users' 3D perception and presence. Future studies may choose different approaches such as investigating the combination of IVR and other representation formats, the adequacy of design solutions developed in immersive environments, or the efficiency of communication in IVR-supported review meetings. Other studies could involve user-centered validation of IVR systems to study workers' behavior in hazardous construction environments, construction accident response, disaster evacuation, heavy equipment operation, or occupants' interactions with building systems – particularly in specialized facilities.

Validity of Instruments

Validation of the PQ and 3DPQ scales should be addressed in future research and would require an in-depth look into three types of validity: construct, criterion, and content validity (Alexandre and Coluci, 2011).

Construct validity refers to the degree to which an instrument is able to measure the construct it purports to be measuring (e.g., presence) rather than a different construct (e.g., attention) (Ginty, 2013). It is given by the theoretical account supporting the construct that a questionnaire is supposed to measure. Thus, to establish construct validity of instruments one may check if the measures of a construct behave according to theory. For this to happen, the variable being measured must have been operationalized to correspond to the true theoretical meaning of a construct. Construct validity can also be verified by comparing the developed questions to existing ones that measure similar constructs in order to check how similar the two sets of questions are. In future studies, one way to demonstrate the construct validity of the adapted PQ and 3DPQ scales is by correlating the outcomes of these questionnaires to those found through previous instruments. Section 4.1.1 provides the motivations for the adoption of the Slater-Usoh-Steed (SUS) PQ (Usoh et al., 2000) as the main reference of the PQ adapted for this research rather than using Witmer and Singer's PQ (1998), which was found to comprise several issues of construct validity.

In turn, criterion validity refers to how closely the data collected agree with an established gold standard (i.e., an external criterion of the construct being measured). If measures match the fold standard, an instrument would be useful for predicting performance or behavior in different situations (past, present, or future). Well-established measurement instruments can act as the criterion against which the criterion validity of the new scale is assessed. Ultimately, criterion validity is consolidated over time as more studies validate the new measurement scale. In the future, the adapted version of the PQ used in this study could be analyzed against measurements using a standard well-accepted PQ such as the SUS (Usoh et al., 2000) – whether a high correlation between the two data sets is detected, the adapted version would have criterion validity (Cohen and Swerdlik, 2005). The major problem with criterion validity checking is the general lack of gold standards, especially when it comes to standards on cognitive responses such as perception and presence in virtual environments. In that case, gold standards may either not provide completely accurate estimates of the true value of a construct or simply not be applicable to a behavioral study.

Lastly, content validity refers to the extent to which questions agree with each other and cover all dimensions of the constructs. In other words, are questions good samples among all possible questions that could have been developed about the construct? (Alexandre and Coluci, 2011). While the adapted PQ comprises questions addressing both dimensions of the presence construct (involvement and immersion), the 3DPQ encompasses various types of questions, each addressing a type of distance estimation. Also, the items of both scales are consistent with each other, as shown in the analyses of presence and 3D perception per question. Therefore, content validity might be partially inferred but should be confirmed upon further examination of scales.

APPENDIX A

Consent Form

Georgia Institute of Technology
CONSENT FORM

Project Title: ***Measuring presence and layout perception in virtual reality environments***
Principal Investigator: Javier Irizarry, Ph.D.

<u>PURPOSE</u>

You are being asked to be a volunteer in a research study that aims at investigating the effectiveness of virtual reality platforms in providing better understanding of three-dimensional computational models ("3D models"). By understanding the extent of potential benefits, more efficient technologies and usage methods could be developed.

<u>PROCEDURES</u>

The experiment consists of three activities. First, you will be asked to fill out a demographic data questionnaire (profession, age, years of experience, etc.), and undergo a paper-based multiple-choice test on spatial ability, which should take approximately 10 minutes. Second, you will look at different modes of an indoor environment representation (the Caddell Building lobby on the Georgia Tech campus): printed two-dimensional drawings, a BIM (Building Information Model) model on a conventional computer screen, and an immersive three-dimensional simulation using a commercial head mounted display (HMD, such as HTC Vive). Simultaneously, you should respond to a set of multiple-choice questions about your layout/spatial perception of the depicted environment. This is followed by another questionnaire about the level of presence (involvement and immersion) you experienced during the perception task. Finally, you will visit the actual space depicted in drawings and 3D models and respond to the very same perception questions. The second and third activities should take about 20 minutes. In total, the experiment should take approximately 30 minutes.

<u>RISKS</u>

This study poses no more than minimal risks to your physical and mental integrity. There is a low risk of experiencing low levels of motion sickness (e.g., nausea, dizziness) while wearing the HMD. The chances of experiencing motion sickness are minimized by allowing you sufficient acquaintance time to get used with the HMD prior to performing the experimental tasks. Please inform the researcher immediately should you feel any discomfort. The researcher will be at your side whenever you are wearing the HMD, and will assist you in taking the HMD off in the event of experiencing motion sickness. If that happens, after taking off the HMD you should remain seated for about two minutes as the discomfort fades away. There is also a near-zero risk of breach of confidentiality. All data collected will be converted into digital format and stored in a password-protected hard drive. Please refer to section "Confidentiality" below for more information.

<u>BENEFITS</u>

You are not likely to receive any direct benefits from the study. However, it will likely benefit the building construction community in general by generating knowledge on the effectiveness of virtual reality technology in architecture and construction practices.

<u>COMPENSATION TO YOU</u>

There is no compensation for participation.

<u>CONFIDENTIALITY</u>

The following procedures will be followed to keep your personal information confidential in this study. Your identity and information will be kept private to the extent required by law. To protect you privacy, your records will be kept under a code number ("participant number") rather than by your name. This means that your records are anonymous. Therefore, it will be impossible to identify you from the data collected. Even the researchers will not have access to your identity from your data. Your data will be stored in locked physical files prior to being turned into digital documents. Once converted into digital format, your data will be stored in a password-protected hard drive, and all hard copy documents with your data will be destroyed. Only research staff is allowed to handle and consult data. Any information that might point to you will not appear when results of this study are presented or published. Once the study ends, your data will be kept stored in digital format, in a password-protected hard drive, for archival purposes for up to five years. After that period it will be erased forever. To make sure that this research is being carried out in the proper way, the Georgia Tech IRB will review the research plan and instruments. The Office of Human Research Protections may also look at study records.

COSTS TO YOU

There are no costs to you, other than your time, for participating in this study.

QUESTIONS ABOUT THE STUDY

Should you have any questions about this study or its procedures, please contact Dr. Javier Irizarry at javier.irizarry@gatech.edu.

QUESTIONS ABOUT YOUR RIGHTS AS A RESEARCH PARTICIPANT

Your participation in this study is voluntary. You do not have to be in this study if for any reason you decide so. You have the right to change your mind and leave the study at any time without penalty. If you decide to withdraw from the study your information will be destroyed. You will be given a copy of this consent form to keep. You do not waive any of your legal rights by agreeing to be in this study. Should you have any questions about your rights as a research subject, please contact Ms. Melanie Clark from the Office of Research Integrity Assurance, at (404) 894-6942, or Ms. Kelly Winn from the same Office, at (404) 385-2175.

CONSENT

I have read this form and received a copy of it. I have had all my questions answered to my satisfaction, and I agree to take part in this study.

____
Participant Name (printed)

____ ______________
Participant Signature Date

____ ______________
Signature of Person Obtaining Consent Date

APPENDIX B

Demographic Questionnaire (DQ)

Participant # : __________

Date: _______/_______/__________

Sequence of presentation: _______ / _______ / PhE

1. **Age**:
   ( ) 18-25
   ( ) 26-33
   ( ) 34-41
   ( ) 42-49
   ( ) 50 and over

2. **Gender**:
   ( ) Female
   ( ) Male
   ( ) Other

3. **Educational level (ongoing)**: ( ) Bachelor ( ) Master ( ) Ph.D.

   Bachelor's degree major: ____

   Current academic major: ____

   (e.g., Architecture, Civil Engineering, Building Construction, Management, Business, etc.)

4. **Occupation (check all that apply)**:
   ( ) Student / RA / TA
   ( ) Intern (in AECO-FM)
   ( ) Faculty (in AECO-FM)
   ( ) Architect
   ( ) Civil Engineer
   ( ) Construction Manager
   ( ) Facility Manager
   ( ) Trade Contractor
   ( ) Other: ____________________

5. **Experience in design review (classroom and industry experience)**:
   ( ) none
   ( ) up to 1 year
   ( ) between 1-5 years
   ( ) between 5-10 years
   ( ) between 10-15 years
   ( ) between 15-20 years
   ( ) over 20 years

6. **Computer usage**:
   ( ) regular use
   ( ) occasional use
   ( ) rare use
   ( ) no use

7. **Experience with 3D virtual environments (BIM, videogames, etc.)**
   ( ) expert
   ( ) intermediate
   ( ) beginner
   ( ) no experience

8. **Have you been to the Caddell Building lobby?**
   ( ) regularly
   ( ) occasionally
   ( ) rarely
   ( ) never

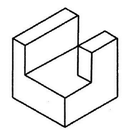
APPENDIX C

Spatial Ability Test

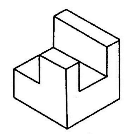
(Revised PSVT:R; Yoon, 2011)

Directions

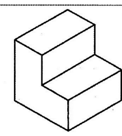
This test consists of 30 questions designed to see how well you can visualize the rotation of three-dimensional objects. Shown below is an example of the type of question included in the second section.



Is Rotated To

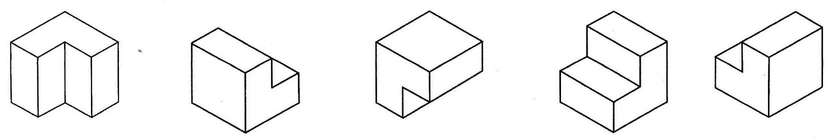


As



Is Rotated To

A B C D E



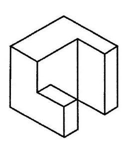
You are to:

1. study how the object in the top line of the question is rotated;
2. picture in your mind what the object shown in the middle line or the question looks like when rotated in exactly the same manner;
3. select from among the five drawings (A, B, C, D, or E) given in the bottom line of the question the one that looks like the object rotated in the correct position.

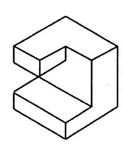
What is the correct answer to the example shown above?

Answers A, B, C, and E are wrong. Only drawing D looks like the object rotated according to the given rotation. Remember that each question has only one correct answer.

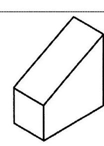
Now look at the next example shown below and try to select the drawing that looks like the object in the correct position when the given rotation is applied.



Is Rotated To

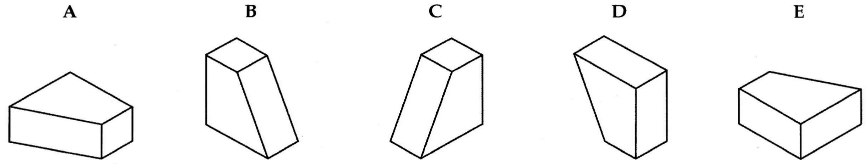


As



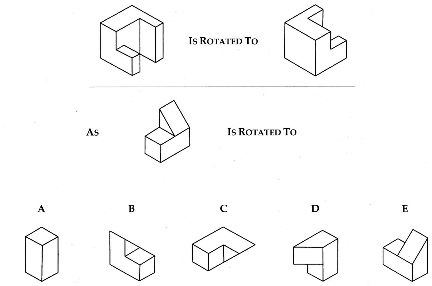
Is Rotated To

A B C D E

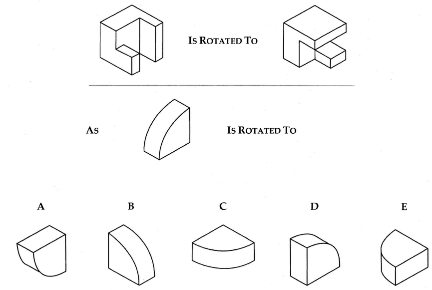


Notice that the given rotation in this example is more complex. The correct answer for this example is B.

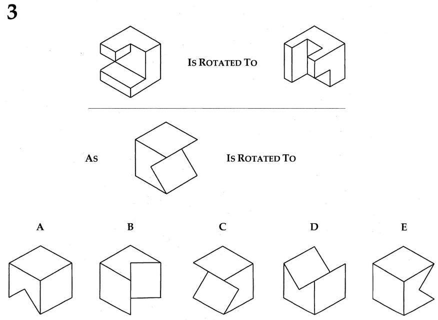
1


Is Rotated To
As
Is Rotated To
A
B
C
D
E

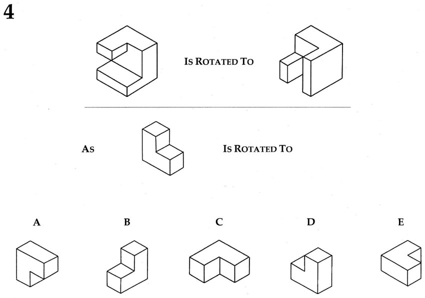
2


Is Rotated To
As
Is Rotated To
A
B
C
D
E

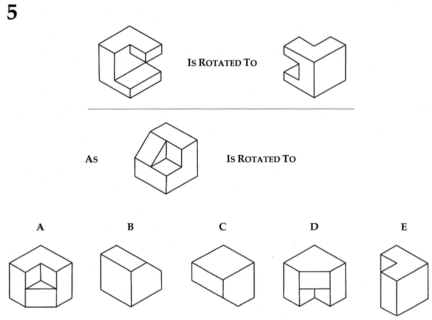
3


Is Rotated To
As
Is Rotated To
A
B
C
D
E

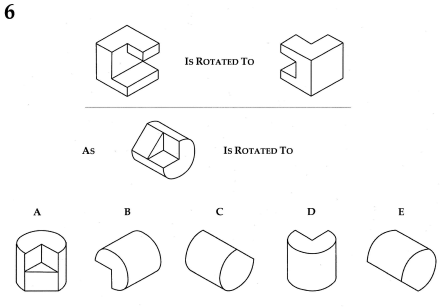
4


Is Rotated To
As
Is Rotated To
A
B
C
D
E

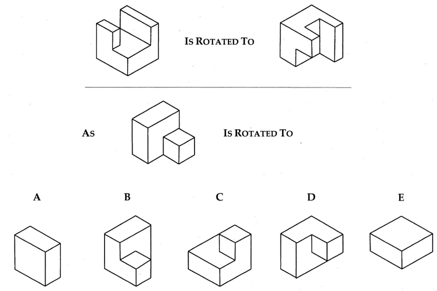
5


Is Rotated To
As
Is Rotated To
A
B
C
D
E

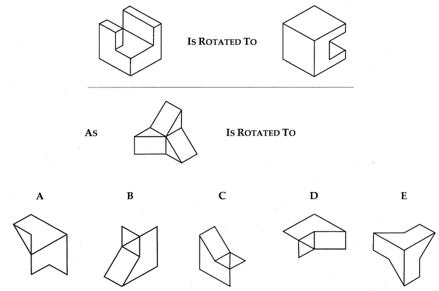
6


Is Rotated To
As
Is Rotated To
A
B
C
D
E

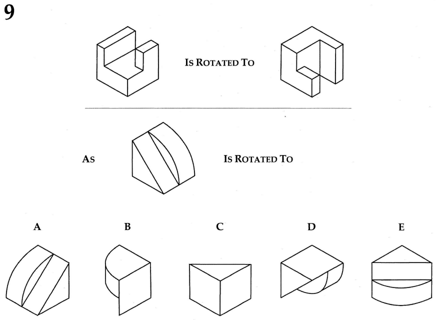
7


Is Rotated To
As
Is Rotated To
A
B
C
D
E

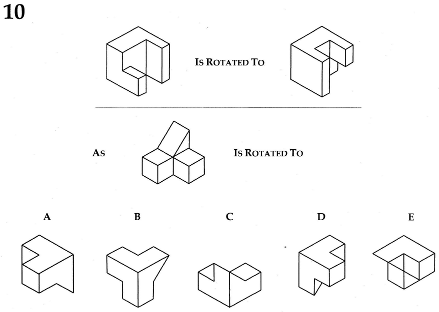
8


Is Rotated To
As
Is Rotated To
A
B
C
D
E

9


Is Rotated To
As
Is Rotated To
A
B
C
D
E

10


Is Rotated To
As
Is Rotated To
A
B
C
D
E

APPENDIX D

3D Perception Questionnaire (3DPQ)

Participant # : __________

Date: _______/_______/__________

Instructions:

- Please answer the questionnaire to the best of your abilities, by choosing a single alternative option. Answer the questionnaire while exploring the environment.
- The environment consists of the entrance hall of the Caddell Building, including the staircase. Do not consider adjacent corridors and upper floor as part of the environment.
- Should you have any questions do not hesitate to ask the researcher.

1. Please stand at the main entrance door. The distance between you and the flex space door is: *(fixed position)*

( ) up to 9'10" (3 m)
( ) up to 16'5" (5 m)
( ) up to 23' (7 m)
( ) up to 29'6" (9 m)
( ) up to 36'1" (11 m)
( ) I can not evaluate

2. Standing at the main entrance door, the distance between you and the drinking fountain to your left is: *(fixed position)*

( ) up to 9'10" (3 m)
( ) up to 16'5" (5 m)
( ) up to 23' (7 m)
( ) up to 29'6" (9 m)
( ) up to 36'1" (11 m)
( ) I can not evaluate

3. Please stand at the staircase end, facing the steps. The distance between you and the wall in front of you is: *(fixed position)*

( ) up to 9'10" (3 m)
( ) up to 16'5" (5 m)
( ) up to 23' (7 m)
( ) up to 29'6" (9 m)
( ) up to 36'1" (11 m)
( ) I can not evaluate

4. Standing at the staircase end, facing the glass curtain. There are two panels in front of you – the glass curtain, and another wall to the right with a vertical window. How distant is the glass curtain from the wall? *(fixed position)*

( ) up to 3'3" (1 m)
( ) up to 6'7" (2 m)
( ) up to 9'10" (3 m)
( ) up to 13'2" (4 m)
( ) up to 16'5" (5 m)
( ) I can not evaluate

5. Please stand at the flex space door, facing the main entrance door. There are two panels in front of you – the entrance door, and a wall to the right with a TV display. How distant is the wall from the entrance door? *(fixed position)*

( ) up to 3'3" (1 m)
( ) up to 6'7" (2 m)
( ) up to 9'10" (3 m)
( ) up to 13'2" (4 m)
( ) up to 16'5" (5 m)
( ) I can not evaluate

6. Standing at the flex space door, facing the center of the lobby. The maximum number of people standing that the space could accommodate (overcrowded) is: *(fixed position)*

( ) up to 30
( ) up to 40
( ) up to 50
( ) up to 60
( ) up to 70
( ) I can not evaluate

7. The distance between the glass curtain wall and the opposite internal wall (adjacent to the staircase) is: *(exploration allowed)*

(   ) up to 19’8” (6 m)
(   ) up to 26’3” (8 m)
(   ) up to 32’10” (10 m)
(   ) up to 39’4” (12 m)
(   ) up to 45’11” (14 m)
(   ) I can not evaluate

8. The distance between the opposite TV displays is: *(exploration allowed)*

(   ) up to 19’8” (6 m)
(   ) up to 26’3” (8 m)
(   ) up to 32’10” (10 m)
(   ) up to 39’4” (12 m)
(   ) up to 45’11” (14 m)
(   ) I can not evaluate

9. How wider is the glass curtain compared to the wall with a vertical window? *(exploration allowed)*

(   ) up to 3’3” (1 m)
(   ) up to 6’7” (2 m)
(   ) up to 9’10” (3 m)
(   ) up to 13’2” (4 m)
(   ) up to 16’5” (5 m)
(   ) I can not evaluate

10. The maximum distance between the floor and the gypsum board ceilings is: *(exploration allowed)*

(   ) up to 6’7” (2 m)
(   ) up to 9’10” (3 m)
(   ) up to 13’2” (4 m)
(   ) up to 16’5” (5 m)
(   ) up to 19’8” (6 m)
(   ) I can not evaluate

11. The maximum distance between the staircase landing and the top edge of the second floor railing is: *(exploration allowed)*

(   ) up to 6’7” (2 m)
(   ) up to 9’10” (3 m)
(   ) up to 13’2” (4 m)
(   ) up to 16’5” (5 m)
(   ) up to 19’8” (6 m)
(   ) I can not evaluate

12. How taller is the flex space door compared to the elevator door? *(exploration allowed)*

(   ) up to 10” (25 cm) taller
(   ) up to 1’8” (50 cm) taller
(   ) up to 2’6” (75 cm) taller
(   ) up to 3’3” (1 m) taller
(   ) up to 4’1” (1.25 m) taller
(   ) I can not evaluate

APPENDIX E

Presence Questionnaire (PQ)

Participant # : __________

Date: _______/_______/__________

Presentation: __________

Instructions:

- Please have in mind the following definition of *presence* when responding to this questionnaire: ***Presence is the sense of being in the place depicted.***
- Please provide your answers to the following questions by choosing a number on a scale from 1 to 7. Answer the questionnaire after exploring the environment.
- Should you have any questions do not hesitate to ask the researcher.

1. To what extent did you feel present in the lobby considering your presence experiences in the real world?

Not at all	1	2	3	4	5	6	7	A great deal

2. When you think back about your experience, to what extent do you think of the lobby as a place in a way similar to when you remember of other places that you have been today?

Not at all	1	2	3	4	5	6	7	A great deal

3. When you think back about your experience, to what extent do you think of the lobby as somewhere you were at?

Not at all	1	2	3	4	5	6	7	A great deal

4. During the time of the experience, how strong was your sense of being in the lobby rather than being in the experiment room?

Not at all	1	2	3	4	5	6	7	Very strong

5. To what extent did your visual experiences in the lobby seem consistent with your visual experiences in the real world?

Not at all	1	2	3	4	5	6	7	A great deal

6. To what extent did you feel you could grasp an object in the lobby?

Not at all	1	2	3	4	5	6	7	A great deal

7. If the lobby ceiling had started to collapse, what would have been the probability of you dodging in an attempt to not getting hit by falling parts?

Not at all	1	2	3	4	5	6	7	Very likely

8. To what extent did you feel like exploring the rest of the environment (second floor, corridors, etc.)?

Not at all	1	2	3	4	5	6	7	A great deal

9. Were there times during the experience when the lobby was the reality for you?

Not at all	1	2	3	4	5	6	7	Almost all times

10. Were you involved in the experience to the extent that you lost track of time?

Not at all	1	2	3	4	5	6	7	A great deal

11. To what extent have you experienced motion sickness (nausea, dizziness)?

Not at all	1	2	3	4	5	6	7	A great deal

APPENDIX F

Research Protocol

Georgia Tech Office of Research Integrity Assurance

Protocol Number: H18334
Funding Agency: n/a
Review Type: Exempt, Category 2
Title: Measuring presence and layout perception in virtual reality environments
Number of Subjects: 60

September 11, 2018

Javier Irizarry
Architecture
javier.irizarry@gatech.edu

Dear Dr. Irizarry:

The Institutional Review Board (IRB) has carefully considered the referenced protocol. Your approval is effective as of **09/11/2018**. The proposed procedures and affiliated documents are exempt from further review by the Georgia Tech Institutional Review Board.

Minimal risk research qualified for exemption status under 45 CFR 46 101b.2.

Thank you for allowing us the opportunity to review your plans. If any complaints or other evidence of risk should occur, or if there is a significant change in the plans, the IRB must be notified.

For your reference, detailed PI responsibilities are included following this letter. If you have any questions concerning this approval or regulations governing human subject activities, please contact me at 404.894.6944.

Sincerely,

Carolyn Sims, MPA, CIP
Office of Research Integrity Assurance
Georgia Institute of Technology

cc: Barbara Henry, IRB Chair

REFERENCES

Adi, M., and Roberts, D. (2014). Using Virtual Environments to Test the Effects of Lifelike Architecture on People. *Technologies of Inclusive Well-Being*, 536, 261–285. DOI: 10.1007/978-3-642-45432-5_13

Al-Hammad, A., Assaf, S., and Al-Shihah, M. (1997). The effect of faulty design on building maintenance. *J. Qual. Mainten. Eng.*, 3(1), 29–39.

Alexandre, N., and Coluci, M. (2011). Content validity in the development and adaptation processes of measurement instruments. *Ciência & Saúde Coletiva,* 16(7), 3061–3068. DOI: 10.1590/S1413-81232011000800006

Bafna, S., and Chambers, E. C. (2014). The influence of spatial organization of the home on inhabitant activity. *A/Z: ITU Journal of Faculty of Architecture*, 11(2), 31–46.

Bassanino, M., Wu, K., Yao, J., Khosrowshahi, F., Fernando, T., and Skjaerbaek, J. (2010). The Impact of Immersive Virtual Reality on Visualization for a Design Review in Construction. *Proceedings of the 14th International Conference on Information Visualization*, 585–589. DOI: 10.1109/IV.2010.85

BCG – Boston Consulting Group. (2018). Available at: http://www3.weforum.org/docs/Future_Scenarios_Implications_Industry_report_2018.pdf

Becerik-Gerber, B., Jazizadeh, F., Li, N., and Calis, G. (2012). Application Areas and Data Requirements for BIM-Enabled Facilities Management. *Journal of Construction Engineering and Management*, 138(3), 431–442. DOI: 10.1061/(ASCE)CO.1943-7862.0000433

Berg, L. P., and Vance, J. M. (2016). An Industry Case Study: Investigating Early Design Decision Making in Virtual Reality. *Journal of Computing and Information Science in Engineering*, 17(1), 011001–011001-7. DOI: 10.1115/1.4034267

Berg, L. P., and Vance, J. M. (2016b). Industry use of virtual reality in product design and manufacturing: a survey. *Virtual Reality*, 21(1), 1–17. DOI: 10.1007/s10055-016-0293-9

Bertol, D. (1997). *Designing Digital Space: An Architect's Guide To Virtual Reality*. John Wiley & Sons, New York. ISBN: 0-471-14662-5

Boton, C. (2018). Supporting constructability analysis meetings with Immersive Virtual Reality-based collaborative BIM 4D simulation. *Automation in Construction*, 96, 1–15. DOI: 10.1016/j.autcon.2018.08.020

Brookes, J., Warburton, M., Alghadier, M., Mon-Williams, M., and Mushtaq, F. (2019). Studying human behavior with virtual reality: The Unity Experiment Framework. *Behavior Research Methods*, 1–9. DOI: 10.3758/s13428-019-01242-0

Brooks, K. R. (2017). Depth Perception and the History of Three-Dimensional Art: Who Produced the First Stereoscopic Images? *I-Perception*, 1–22. DOI: 10.1177/2041669516680114

Calderon-Hernandez, C., Paes, D., Irizarry, J., and Brioso, X. (2019). Comparing Virtual Reality and 2-Dimensional Drawings for the Visualization of a Construction Project. *ASCE International Conference on Computing in Civil Engineering (i3CE 2019)*, Atlanta, 25–31. DOI: 10.1061/9780784482421.003

Castronovo, F., Nikolic, D., Liu, Y., and Messner, J. I. (2013). An evaluation of Immersive Virtual Reality systems for design reviews. *Proceedings of the 13th International Conference on Construction Applications of Virtual Reality (CONVR 2013)*, London, UK, 22–29. ISBN: 978-0-9927161-0-3

Castronovo, F., Oprean, D., Liu, T., and Messner, J. (2017). Application of Immersive Virtual Reality Systems in an Interdisciplinary Design Studio Course. *Proceedings of the Joint Conference on Computing in Construction (JC3)*, Heraklion, Greece, 337–345. DOI: 10.24928/JC3-2017/0006

Chambers, E. C., Bafna, S., and Machry, H. (2018). The Association Between Apartment Layout and Depressive Symptomology among Hispanic/Latino Residents in Low-Income Housing: the AHOME Study. *Journal of Urban Health*, 95(1), 51–60. DOI: 10.1007/s11524-017-0216-4

Chandrasegaran, S. K., Ramani, K., Sriram, R. D., Horváth, I., Bernard, A., Harik, R. F., and Gao, W. (2013). The evolution, challenges, and future of knowledge representation in product design systems. *Computer Aided Design*, 45(2), 204–228. DOI: 10.1016/j.cad.2012.08.006

Charness, G., Gneezyb, U., and Kuhn, M. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, 81, 1–8. DOI: 10.1016/j.jebo.2011.08.009

Chen, Q., Weidner, R., Vossel, S., Weiss, P. H., and Fink, G. R. (2012). Neural Mechanisms of Attentional Reorienting in Three-Dimensional Space. *Journal of Neuroscience*, 32(39), 13352–13362. DOI: 10.1523/JNEUROSCI.1772-12.2012

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. 2nd Edition. Lawrence Earlbaum Associates, Hillsdale, NJ.

Cohen, R. J., and Swerdlik, M. E. (2005). *Psychological testing and assessment: An introduction to tests and measurement* (6th Edition). New York: McGraw-Hill.

Cruz-Neira, C., Sandin, D. J., and DeFanti, T. A. (1993). Surround-screen projection-based virtual reality: the design and implementation of the CAVE. *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*, Anaheim, 135–142. DOI: 10.1145/166117.166134

Cutting, J. E., and Vishton, P. M. (1995). Perceiving Layout and Knowing Distances: The Integration, Relative Potency, and Contextual Use of Different Information about Depth. *Information Potency and Spatial Layout*, 69–117.

Digi-Capital. (2019). Available at: https://www.digi-capital.com/news/2019/01/for-ar-vr-2-0-to-live-ar-vr-1-0-must-die/

Dunston, P. S., Arns, L. I., and Mcglothlin, J. D. (2011). An immersive virtual reality mock-up for design review of hospital patient rooms. *Collaborative Design in Virtual Environments, Intelligent Systems, Control and Automation: Science and Engineering*, 167–176. DOI:

10.1007/978-94-007-0605-7_15

Eastman, C., Teicholz, P., Sacks, R., and Liston, K. (2008). *BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers, and Contractors*. John Wiley & Sons, Hoboken. ISBN: 978-0-470-18528-5

England, R. D., Edgar, G. K., and Carr, K. T. (1992). Human factors in virtual worlds. 2. Enabling interaction and performance. *Proceeding of the IEE Colloquium on Using Virtual Worlds*, London.

Faas, D., Bao, Q., Frey, D. D., and Yang, M. C. (2014). The influence of immersion and presence in early stage engineering designing and building. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 28(02), 139–151. DOI: 10.1017/S0890060414000055

Fernando, T., Wu, K. C., and Bassanino, M. (2013). Designing a novel virtual collaborative environment to support collaboration in design review meetings. *Journal of Information Technology in Construction (ITcon)*, 18, 372–396. ISSN: 1874-4753

Florio, W. (2011). An analysis of the design process under the cognition theory: seven difficulties in the atelier. *Arquiteturarevista*, 7(2), 161–171. DOI: 10.4013/arq.2011.72.06

Gao, X., and Pishdad-Bozorgi, P. (2018). Past, Present, and Future of BIM-Enabled Facilities Operation and Maintenance. *Proceedings of the CRC 2018 – Construction Research Congress*, Louisiana State University, April 2-5, 51–61. DOI: 10.1061/9780784481295.006

Gheisari, M. (2013). An Ambient Intelligent Environment for Accessing Building Information in Facility Management Operations; A healthcare Facility Scenario. Ph.D. Dissertations, Georgia Institute of Technology.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin. ISBN 10: 1848725779

Gifford, R. (2002). *Environmental Psychology: Principles and Practice*. Optimal Books, Colville. ISBN: 0968854303

Ginty, A. T. (2013). *Construct Validity*. In: Gellman M. D., Turner J. R. (Eds.), Encyclopedia of Behavioral Medicine. New York: Springer. DOI: 10.1007/978-1-4419-1005-9

Gooch, A. A., and Willemsen, P. (2002). Evaluating Space Perception in NPR Immersive Environments. *Proceedings of the 2nd International Symposium on Non-Photorealistic Animation and Rendering*, 105–110. DOI 10.1145/508530.508549

Green, P., and MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493–498. DOI: 10.1111/2041-210X.12504

Hammer, Ø., Harper, D. A. T., Ryan, P. D. (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, 4(1).

Henry, D., and Furness, T. (1993). Spatial perception in virtual environments: Evaluating an architectural application. *Proceedings of the IEEE Virtual Reality Annual International Symposium (VRAIS '93)*, Seattle, 33–40. DOI: 10.1109/VRAIS.1993.380801

Heydarian, A., Carneiro, J. P., Gerber, D., and Becerik-Gerber, B. (2015a). Immersive virtual environments, understanding the impact of design features and occupant choice upon lighting for building performance. *Building and Environment*, 89, 217–228. DOI: 10.1016/j.buildenv.2015.02.038

Heydarian, A., Carneiro, J. P., Gerber, D., Becerik-Gerber, B., Hayes, T., and Wood, W. (2014). Immersive Virtual Environments: Experiments on Impacting Design and Human Building Interaction. *Proceedings of the 19th International Conference on Computer-Aided Building Design Research in Asia (CAADRIA)*, Kyoto, 729–738. ISBN: 978-988-19026-5-8

Heydarian, A., Carneiro, J. P., Gerber, D., Becerik-Gerber, B., Hayes, T., and Wood, W. (2015b). Immersive virtual environments versus physical built environments: A benchmarking study for building design and user-built environment explorations. *Automation in Construction*, 54, 116–126. DOI: 10.1016/j.autcon.2015.03.020

Heydarian, A., Carneiro, J. P., Pantazis, E., Gerber, D., and Becerik-Gerber, B. (2015c). Default Conditions: A Reason for Design to Integrate Human Factors. *Proceedings of the First International Symposium on Sustainable Human-Building Ecosystems*, Pittsburgh, 54–62. DOI: 10.1061/9780784479681.006

Higuera-Trujillo, J. L., Lopez-Tarruella Maldonado, J., and Millan, C. L. (2017). Psychological and physiological human responses to simulated and real environments: A comparison between Photographs, 360° Panoramas, and Virtual Reality. *Applied Ergonomics*, 65, 398–409. DOI: 10.1016/j.apergo.2017.05.006

Hochberg, J. (1998). *Gestalt Theory and Its Legacy*. In: J. Hochberg (Ed.), Perception and Cognition at Century's End (2nd Edition, 253–306). San Diego London Boston New York Sydney Tokyo Toronto: Academic Press.

Hoffman, D. (1998). *Visual Intelligence: How We Create What We See*. New York: W. W. Norton.

Hoffman, D., and Singh, M. (2006). *Vision: Form Perception*. In: Nadel, Lynn (Ed.), Encyclopedia of Cognitive Science.

Hubona, G. S., Shirah, G. W., and Fout, D. G. (1997). The effects of motion and stereopsis on three-dimensional visualization. *International Journal of Human-Computer Studies*, 47, 609–627.

Hubona, G. S., Wheeler, P. N., Shirah, G. W., and Brandt, M. (1999). The relative contributions of stereo, lighting, and background scenes in promoting 3D depth visualization. ACM Transactions. *Computer-Human Interaction*, 6(3), 214–242.

Ibrahim, R., and Rahimian, F. P. (2010). Comparison of CAD and manual sketching tools for teaching building design. *Automation in Construction*, 19(8), 978–987. DOI: 10.1016/j.autcon.2010.09.003

Interrante, V., Anderson, L., and Ries, B. (2006). Distance perception in immersive virtual environments, revisited. *Proceedings of the IEEE Virtual Reality Conference*, 3–10. DOI:

10.1109/VR.2006.52

Interrante, V., Ries, B., Lindquist, J., Kaeding, M., and Anderson, L. (2008). Elucidating factors that can facilitate veridical spatial perception in immersive virtual environments. *Presence*, 17(2), 176–198. DOI: 10.1162/pres.17.2.176

IoT-Analytics. (2019). Available at: https://iot-analytics.com/top-20-industrial-iot-trends-hannover-messe-2019/

Kalawsky, R. S., Bee, S. T., and Nee, S. P. (1999). Human factors evaluation to aid understanding of virtual interfaces. *BT Technology*, 17(1), 128–141.

Kalisperis, L., Muramoto, K., Balakrishnan, B., Nikolic, D., and Zikic, N. (2006). Evaluating relative impact of virtual reality system variables on building design comprehension and presence: a variable-centered approach using fractional factorial experiment. *Proceedings of the 24th Education and Research in Computer Aided Building Design in Europe (eCAADe)*, Volos, Greece, 66–73. ISBN: 0-9541183-5-9

Kalisperis, L., Otto, G., Muramoto, K., Gundrum, J., Masters, R., and Orland, B. (2002). An Affordable Immersive Environment in Beginning Design Studio Education. *Proceedings of the Annual Conference of the Association for Computer Aided Design In Architecture (ACADIA)*, Pomona, 49–56. ISBN: 1-880250-11-X

Kelsey, C. A. (1993). *Detection of visual information*. In: W. R. Hendee, and P. Wells (Eds.), The Perception of Visual Information (30–51). New York: Springer.

Khashe, S., Becerik-Gerber, B., Lucas, G., and Gratch, J. (2018). Persuasive Effects of Immersion in Virtual Environments for Measuring Pro-Environmental Behaviors. *Proceedings of the 35th International Symposium on Automation and Robotics in Construction (ISARC 2018)*, Berlin, Germany, 1205–1211. DOI: 10.22260/ISARC2018/0167

Khemlani, L., Timerman, A., Beene, B., and Kalay, Y. (1997). Semantically Rich Building Representation. *Proceeding of the 16th Association for Computer Aided Design in Architecture Conference (ACADIA)*, Cincinnati, 207–228. ISBN: 1880250063

Khuu, S. K., Gordon, J., Balcomb, K., and Kim, J. (2014). The perception of three-dimensional cast-shadow structure is dependent on visual awareness. *Journal of Vision*, 14(25). DOI: 10.1167/14.3.25

Kim, M. J., Wang, X., Love, P. ED., Li, H., and Kang, SC. (2013). Virtual reality for the built environment: a critical review of recent advances. *Journal of Information Technology in Construction (ITcon)*, 18, 279–304. ISSN: 1874-4753

Kimura, K., Reichert, J. F., Olson, A., Pouya, O. R., Wang, X., Moussavi, Z., and Kelly, D. M. (2017). Orientation in Virtual Reality Does Not Fully Measure Up to the Real-World. *Scientific Reports*, 7(1). DOI: 10.1038/s41598-017-18289-8

Kinsey, B., Towle, E., Hwang, G., O'Brien, E. J., and Bauer, C. F. (2006). The Effect Of Spatial Ability On The Retention Of Students In A College Of Engineering And Physical Science. *Proceedings of the American Society for Engineering Education Annual Conference & Exposition (ASEE 2006)*, Chicago, United States, 11.1279.1–11.1279.12.

Kosslyn, S. M. (1994). *Image and Brain: The Resolution of the Imagery Debate*. MIT Press, Cambridge. ISBN: 0-262-11184-5

Kovac, R. J. (1989). The validation of selected spatial ability tests via correlational assessment and analysis of user-processing strategy. *Educational Research Quarterly*, 13(2), 26–34.

KPMG – Global Construction Survey. (2016). Available at: https://assets.kpmg/content/dam/kpmg/xx/pdf/2016/09/global-construction-survey-2016.pdf

Kuliga, S. F., Thrash, T., Dalton, R., and Hölscher, C. (2015). Virtual reality as an empirical research tool – Exploring user experience in a real building and a corresponding virtual model. *Computers, Environment and Urban Systems*, 54, 363–375. DOI: 10.1016/j.compenvurbsys.2015.09.006

Kumar, S., Hedrick, M., Wiacek, C., and Messner, J. I. (2011). Developing an experienced-based design review application for healthcare facilities using a 3d game engine. *Journal of Information Technology in Construction (ITcon)*, 16, 85–104. ISSN: 1874-4753

Lazar, J., Feng, J. H., and Hochheiser, H. (2017). *Research methods in human-computer interaction*. 2nd Edition. Amsterdam: Morgan Kaufmann.

Leicht, R. M., Kumar, S., Abdelkarim, and Messner, J. (2010). Gaining End User Involvement through Virtual Reality Mock-Ups: A Medical Facility Case Study. *CIB W78 2010 – Applications of IT in the AEC Industry*. http://itc.scix.net/cgi-bin/works/Show?w78-2010-143

Liang, K., and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1), 13–22. DOI: 10.1093/biomet/73.1.13

Lovreglio, R., Gonzalez, V., Feng, Z., Amor, R., Spearpoint, M., Thomas, J., Trotter, M., and Sacks, R. (2018). Prototyping virtual reality serious games for building earthquake preparedness: The Auckland City Hospital case study. *Advanced Engineering Informatics*, 38, 670–682. DOI: 10.1016/j.aei.2018.08.018

Lucas, J. (2018). Immersive VR in the construction classroom to increase student understanding of sequence, assembly, and space of wood frame construction. *Journal of Information Technology in Construction (ITcon)*, 23, 179–194. ISSN: 1874-4753

Maher, M. L. (2011). Designers and Collaborative Virtual Environments. *Collaborative Design in Virtual Environments, Intelligent Systems, Control and Automation: Science and Engineering*, 3–15. DOI: 10.1007/978-94-007-0605-7_1

Majumdar, T., Fischer, M. A., and Schwegler, B. R. (2006). Conceptual Design Review with a Virtual Reality Mock-Up Model. *Proceedings of the Joint International Conference on Computing and Decision Making in Civil and Building Engineering*, Montréal, 2902–2911. ISBN: 2-921145-58-8

Maldovan, K., and Messner, J. (2006). Determining the effects of immersive environments on decision making in the AEC Industry. *Proceedings of the Joint International Conference on Computing and Decision Making in Civil and Building Engineering*, Montréal, 2762–2770. ISBN: 2-921145-58-8

Maldovan, K., Messner, J., and Faddoul, M. (2006). Framework for reviewing mockups in an immersive environment. *Proceedings of the 6th International Conference on Construction Applications of Virtual Reality (CONVR 2006)*, Orlando. Available at http://itc.scix.net/pdfs/6d87.content.04272.pdf

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman, New York. ISBN: 0716715678

McGinn, C. (2004). *Mindsight: Image, Dream, Meaning*. Harvard University Press.

Norcia, A. M., and Gerhard, H. E. (2015). Development of Three-Dimensional Perception in Human Infants. *Annual Review of Vision Science*, 1, 569–94. DOI: 10.1146/annurev-vision-082114-035835

Nowak, K. L., Krcmar, M., and Farrar K. M. (2008). The causes and consequences of presence: Considering the influence of violent video games on presence and aggression. *Presence*, 17(3), 256–268. DOI: 10.1162/pres.17.3.256

Okamoto, J. (2014). *Percepção Ambiental e Comportamento: Visão Holística na Arquitetura e na Comunicação*. Mackenzie, São Paulo. ISBN: 8582930364

Okeil, A. (2010). Hybrid design environments: immersive and non-immersive building design. *Journal of Information Technology in Construction (ITcon)*, 15, 202–216. ISSN: 1874-4753

Oren, M., Carlson, P., Gilbert, S., and Vance, J. M. (2012). Puzzle Assembly Training: Real World vs. Virtual Environment. *Proceedings of the IEEE Virtual Reality Conference*, Costa Mesa, United States, 27–30. DOI: 10.1109/VR.2012.6180873

Otto, G., Kalisperis, L., Gundrum, J., Muramoto, K., Burris, G., Masters, R., Slobounov, E., Heilman, J., and Agarwala, V. (2003). The VR-Desktop: an Accessible Approach to VR Environments in Teaching and Research. *International Journal of Building Computing*, 1(2), 233–246. DOI: 10.1260/147807703771799201

Owen, R., Amor, R., Palmer, M., Dickinson, J., Tatum, C. B., Abdul Samad, K., Prins, M., Kiviniemi, A., and East, B. (2010). Challenges for Integrated Design and Delivery Solutions. *Building Engineering and Design Management*, 6(4), 232–240. DOI: 10.3763/aedm.2010.IDDS1

Ozcelik, G., and Becerik-Gerber, B. (2018). Benchmarking thermoception in virtual environments to physical environments for understanding human-building interactions. *Advanced Engineering Informatics*, 36, 254–263. DOI: 10.1016/j.aei.2018.04.008

Paes, D., and Irizarry, J. (2016). Virtual Reality Technology Applied in the Building Design Process: Considerations on Human Factors and Cognitive Processes. *Advances in Ergonomics in Design*, 3–15. DOI: 10.1007/978-3-319-41983-1_1

Paes, D., and Irizarry, J. (2018). A Usability Study of an Immersive Virtual Reality Platform for Building Design Review: Considerations on Human Factors and User Interface. *Proceedings of the CRC 2018 – Construction Research Congress*, Louisiana State University, April 2-5. DOI: 10.1061/9780784481264.041

Paes, D., and Irizarry, J. (2019). The Relevance of Visual Cues in Immersive Environments: Does Pictorial Realism Matter? *ASCE International Conference on Computing in Civil Engineering* *(i3CE 2019)*, Atlanta, 25–31. DOI: 10.1061/9780784482421.004

Paes, D., Arantes, E., and Irizarry, J. (2017). Immersive environment for improving the understanding of architectural 3D models: Comparing user spatial perception between immersive and traditional virtual reality systems. *Automation in Construction*, 84, 292–303. DOI: 10.1016/j.autcon.2017.09.016

Palmer, S. E. (2003). *Visual Perception of Objects*. Handbook of Psychology: Experimental Psychology, Three:7:177–211. John Wiley & Sons. DOI: 10.1002/0471264385.wei0407

Pashler, H. E. (1998). *The psychology of attention*. MIT Press, Cambridge. ISBN: 0-262-16165-6

Patrick, E., Cosgrove, D., Slavkovic, A., Rode, J. A., Verratti, T., and Chiselko, G. (2000). Using a large projection screen as an alternative to head-mounted displays for virtual environments. *Proceedings of the Conference on Human Factors in Computing Systems (SIGCHI)*, Hague, 478–485. DOI: 10.1145/332040.332479

Portman, M. E., Natapov, A., and Fisher-Gewirtzman, D. (2015). To go where no man has gone before: Virtual reality in architecture, landscape architecture and environmental planning. *Computers, Environment and Urban Systems*, 54, 376–384. DOI: 10.1016/j.compenvurbsys.2015.05.001

Pourhoseingholi, M. A., Baghestani, A. R., and Vahedi, M. (2012). How to control confounding effects by statistical analysis. *Gastroenterology and hepatology from bed to bench*, 5(2), 79–83.

Pujoni, D. G. F. (2019). *Análises Estatísticas Passo a Passo (Step-by-Step Statistical Analyses)*. 1st Edition. Belo Horizonte. Available at: http://scimus.com.br/

Renner, R. S., Velichkovsky, B. M., and Helmert, J. R. (2013). The perception of egocentric distances in virtual environments – A review. *ACM Computing Surveys*, 46(2). DOI: 10.1145/2543581.2543590

Rosenthal, R., and Rosnow, R. (2008). *Essentials of Behavioral Research: Methods and Data Analysis*, 3rd ed. McGraw Hill, Boston, MA.

Ruschel, R., Fracaroli, F., and Silva, T. (2005). Realidade Virtual Exploratória utilizando-se o Quicktime VR: um estudo sobre a percepção do ambiente simulado para avaliação de projeto. *Proceedings of the 2nd Seminar on Information Technology and Communication in Civil Construction (TIC 2005)*, EPUSP, São Paulo, 1–10.

Sacks, R., Perlman, A., and Barak, R. (2013). Construction safety training using immersive virtual reality. *Construction Management and Economics*, 31(9), 1005–1017. DOI: 10.1080/01446193.2013.828844

Schnabel, M. A., and Kvan, T. (2003). Spatial Understanding in Immersive Virtual Environments. *International Journal of Building Computing*, 1(4), 435–448. DOI: 10.1260/147807703773633455

SEP – Startup Europe Partnership. (2018). Research Highlights: Digital Construction and Industry 4.0. Available at: https://startupeuropepartnership.eu/wp-content/uploads/2018/03/SEP-ResearchHighlights_Industry4.0.pdf

Shop Architects. (2019). Available at: https://www.shoparc.com

Sinai, M. J., Krebs, W. K., Darken, R. P., Rowland, J. H., and McCarley, J. S. (1999). Egocentric distance perception in a virtual environment using a perceptual matching task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 43(22), 1256–1260. DOI: 10.1177/154193129904302219

Sopher, H., Fisher-Gewirtzman, D., and Kalay, Y. (2018). Use of Immersive Virtual Environment in the Design Studio: An Assessment Model. *Proceedings of the Education and Research in Computer Aided Architectural Design in Europe Conference (eCAADe)*, Lodz, Poland, 853–862.

Sopher, H., Kalay, Y., and Fisher-Gewirtzman, D. (2017). Why Immersive? Using an Immersive Virtual Environment in Architectural Education. *Proceedings of the Education and Research in Computer Aided Architectural Design in Europe Conference (eCAADe)*, Rome, Italy, 313–322.

Stanney, K. M., Mourant, R. R., and Kennedy, R. S. (1998). Human factors issues in virtual environments: A review of the literature. *Presence*, 7(4), 327–351. DOI: 10.1162/105474698565767

Steuer, J. (1992). Defining Virtual Reality: Dimensions Determining Telepresence. *Journal of Communication*, 42(4), 73–93. DOI: 10.1111/j.1460-2466.1992.tb00812.x

Szumilas M. (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 19(3), 227–229.

Thompson, V., and Campbell, J. (2004). A power struggle: between- vs. within-subjects designs in deductive reasoning research. *Psychologia*, 47, 277–296.

Thompson, W. B., Willemsen, P., Gooch, A. A., Creem-Regehr, S. H., Loomis, J. M., and Beall, A. C. (2004). Does the quality of the computer graphics matter when judging distances in visually immersive environments? *Presence*, 13(5), 560–571. DOI: 10.1162/1054746042545292

Tizani, W. (2011). Collaborative Design in Virtual Environments at Conceptual Stage. *Collaborative Design in Virtual Environments, Intelligent Systems, Control and Automation: Science and Engineering*, 67–76. DOI: 10.1007/978-94-007-0605-7_6

Wang, M. (2014). Generalized estimating equations in longitudinal data analysis: a review and recent developments. *Advances in Statistics*. DOI: 10.1155/2014/303728

Wang, Y., Wang, X., Wang, J., Yung, P., and Jun, G. (2013). Engagement of Facilities Management in Design Stage through BIM: Framework and a Case Study. *Advances in Civil Engineering*, vol. 2013, 8 pages. DOI: 10.1155/2013/189105

Wann, J., and Mon-Williams, M. (1996). What does virtual reality NEED?: human factors issues in the design of three-dimensional computer environments. *International Journal of Human-Computer Studies*, 44(6), 829–847. DOI: 10.1006/ijhc.1996.0035

Witmer, B. G., and Sadowski, W. J. (1998). Nonvisually guided locomotion to a previously viewed target in real and virtual environments. *Human Factors*, 40(3), 478–488. DOI: 10.1518/001872098779591340

Witmer, B. G., and Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence*, 7(3), 225–240. DOI: 10.1162/105474698565686

Xie, H., Tudoreanu, M. E., and Shi, W. (2006). Development of a Virtual Reality Safety-Training System for Construction Workers. *Proceedings of the 6th International Conference on Construction Applications of Virtual Reality (CONVR 2006)*, Orlando. Available at http://itc.scix.net/data/works/att/ff9b.content.00092.pdf

Yabuki, N. (2011). Impact of Collaborative Virtual Environments on Design Process. *Collaborative Design in Virtual Environments, Intelligent Systems, Control and Automation: Science and Engineering*, 103–110. DOI: 10.1007/978-94-007-0605-7_9

Yang, X., and Ergan, S. (2015). Design and Evaluation of an Integrated Visualization Platform to Support Corrective Maintenance of HVAC Problem–Related Work Orders. *Journal of Computing in Civil Engineering*, 30(3). DOI: 10.1061/(ASCE)CP.1943-5487.0000510

Yoon, S. Y. (2011). Psychometric properties of the Revised Purdue Spatial Visualization Tests: Visualization of Rotations (The Revised PSVT:R). PhD Dissertation. Purdue University, West Lafayette, USA.

Ziemer, C., Plumert, J. M., Cremer, J. F., and Kearney, J. K. (2009). Estimating distance in real and virtual environments: Does order make a difference? *Attention, Perception, & Psychophysics*, 71(5), 1095–1106. DOI: 10.3758/APP.71.5.1096